目　录

第一章　线性空间与线性变换

本章扼要概述线性空间与线性变换的基本概念和基本理论．这些材料是学习矩阵分析及其应用的入门知识．对于线性代数基础较好的读者，有些部分粗看一下就可以了．

§1　线性空间的概念

人们谈论问题，往往都是就一定"范围"来说的，脱离了这个"范围"，就难以讲得清了，甚至只能在某个"范围"内才能提出或研究某种问题．明白了这一点，就较易理解我们引入数域及线性空间的目的了．

我们知道，由所有有理数组成的集合具有这样的性质：这集合中任意两数的和、差、积、商(除数不为零)仍是这集合中的数．这个集合我们用 Q 来表示．类似地，由所有实数构成的集合 R，以及由所有复数构成的集合 C 也都具有这一性质．这三个集合的包含关系为：

$$Q \subset R \subset C .$$

因此，我们说"一个复数"时，自然包括这个数可能是有理数或实数这两个特殊情况在内．

在引入线性空间这一重要概念之前，我们首先要给出数域的概念．

如果复数的一个非空集合 P 含有非零的数，且其中任意两数的和、差、积、商(除数不为零)仍属于该集合，则称数集 P 为一个

数域。于是上述的集合 Q,R,C 都是数域，分别称为有理数域、实数域及复数域．又如集合

$$Q(\sqrt{2})=\{a+b\sqrt{2}\mid a,b\in Q\}$$

也构成一数域，请读者加以验证．但是由所有整数组成的集合 Z 是不构成数域的。

数域有一个简单性质，即所有的数域都包含有理数域作为它的一部分。特别，每个数域都包含整数 0 和 1．现在我们可以给出线性空间的定义了。

定义 1 设 V 是一非空集合，P 是一数域。如果：

1) 在集合 V 上定义了一个二元运算(通常称为*加法*)，即是说，V 中任意两个元素 x,y 经过这个运算后所得到的结果，仍是集合 V 中一个唯一确定的元素，这元素称为 x 与 y 的*和*，并记作 $x+y$；

2) 在数域 P 与集合 V 的元素之间还定义了一种运算，叫做*数量乘法*，即是说，对于 P 中任一数 λ 与 V 中任一元素 x，经这一运算后所得结果仍为 V 中一个唯一确定的元素，称为 λ 与 x 的*数量乘积*，记作 λx；

3) 上述两个运算满足下列八条规则：

(1) 对任意 $x,y\in V$，$x+y=y+x$；

(2) 对任意 $x,y,z\in V$，$(x+y)+z=x+(y+z)$；

(3) V 中存在一个*零元素*，记作 θ，对任一 $x\in V$，都有 $x+\theta=x$；

(4) 任一 $x\in V$，都有 $y\in V$，使得 $x+y=\theta$，元素 y 称为 x 的*负元素*，记作 $-x$；

(5) 对任一 $x\in V$，都有 $1x=x$；

$$\left.\begin{array}{l}(6)\ \lambda(\mu x)=(\lambda\mu)x\\(7)\ (\lambda+\mu)x=\lambda x+\mu x\\(8)\ \lambda(x+y)=\lambda x+\lambda y\end{array}\right\}\text{对任何 }\lambda,\mu\in P,\quad x,y\in V$$

则集合 V 称为数域 P 上的线性空间或向量空间。V 中的元素常称为向量。V 中的零元素称为零向量。当 P 是实数域时，V 叫实线性空间；当 P 是复数域时，V 叫复线性空间。数域 P 上的线性空间有时简称为线性空间。

由定义可以证明：

线性空间 V 中的零向量是唯一的；V 中每个元素 x 的负元素也是唯一的；并且有

$$0x=\theta,\quad \lambda\theta=\theta,\quad (-1)x=-x,$$

这里 $\lambda\in P, x\in V$，又 V 中元素的减法可以定义为（对任何 $x,y\in V$）：

$$x-y=x+(-y).$$

线性空间的例子：

例 1　若 P 是数域，V 是分量属于 P 的 n 元有序数组的集合

$$V=\{(\xi_1,\xi_2,\cdots,\xi_n)\mid \forall\ \xi_i\in P\},$$

若对 V 中任二元素

$$x=(\xi_1,\xi_2,\cdots,\xi_n),\quad y=(\eta_1,\eta_2,\cdots,\eta_n)$$

及每个 $\lambda\in P$（记作 $\forall\ \lambda\in P$），定义加法及数量乘法为：

$$x+y=(\xi_1+\eta_1,\xi_2+\eta_2,\cdots,\xi_n+\eta_n),$$
$$\lambda x=(\lambda\xi_1,\lambda\xi_2,\cdots,\lambda\xi_n),$$

则容易验证集合 V 构成数域 P 上的线性空间，这个线性空间记为 P^n。

例 2　所有元素属于数域 P 的 $m\times n$ 矩阵组成的集合，按通常定义的矩阵加法及数与矩阵的数量乘法，也构成数域 P 上的一个线性空间，并把它记为 $P^{m\times n}$。

例 3　若 n 为正整数，P 是数域，则系数属于 P 而未定元为 t 的所有次数小于 n 的多项式的集合，按通常多项式的加法及数与多项式的乘法，这个集合连同零多项式在内亦构成数域 P 上的线性空间。我们用 $P[t]_n$ 代表这个空间。若把"次数小于 n 的"这一限

制取消，则也得到一个线性空间，并记为 $P[t]$。

例 4 所有定义在区间 $[a,b]$ $(a\leqslant t\leqslant b)$ 上的实值连续函数构成的集合，按照函数的加法及数与函数的乘法，显然构成实数域上一个线性空间。

在讨论线性空间的问题中，下面几个基本概念是必须熟知的。

定义 2 设 V 是数域 P 上的线性空间，$x,y,\cdots,v$ 是 V 的一组向量，如果 P 中有一组不全为零的数 $\alpha,\beta,\cdots,\delta$，使得

$$\alpha x+\beta y+\cdots+\delta v=\theta, \qquad (1)$$

则称向量 $x,y,\cdots,v$ 线性相关；若等式(1)仅当 $\alpha=\beta=\cdots=\delta=0$ 时才能成立，则称这组向量是线性无关的。

由这定义得知，如果向量 $x,y,\cdots,v$ 线性相关，则使得(1)式成立的数 $\alpha,\beta,\cdots,\delta$ 中至少有一个不等于零，比如 $\alpha\neq 0$，则有

$$x=-\frac{\beta}{\alpha}y-\cdots-\frac{\delta}{\alpha}v,$$

这时我们说向量 x 是向量 $y,\cdots,v$ 的线性组合，或者说，向量 x 可由向量 $y,\cdots,v$ 线性表出(示)。

一般地说，一组向量(含有限个向量)线性相关时，则其中至少有一个向量可由组中其他向量线性表出；反过来，如果这组向量具有这一性质，则这组向量必定线性相关。但不难推知，线性无关的一组向量，其任一向量都不可能由组中其余向量线性表出。

定义 3 设 V 是数域 P 上的线性空间。如果 V 中有 n 个线性无关的向量，而无更多个数的线性无关向量，则称 V 是数域 P 上的 n 维线性空间。

若线性空间 V 内能求得任意个数的线性无关向量，则 V 便称为无限维线性空间。本书主要讨论有限维线性空间。

在 n 维线性空间中，其任意的 n 个线性无关向量都称为它的一组基。由线性空间维数定义可知，在有限维线性空间中，基是存在的，但不唯一，因为当维数是 n 时，空间里的任何 n 个线性无关

向量都可以作为它的一组基。然而，我们却有下述定理。

定理 1 设 V 是数域 P 上的 n 维线性空间，$e_1, e_2, \cdots, e_n$ 是 V 的一组基。则 V 中任一向量 x 都可以表示为这组基的线性组合，且表示式是唯一的。

证明 因 V 中 $n+1$ 个向量 $x, e_1, e_2, \cdots, e_n$ 线性相关，故 P 中存在 $n+1$ 个不全为零的数 $a_0, a_1, \cdots, a_n$，使得

$$a_0 x + a_1 e_1 + a_2 e_2 + \cdots + a_n e_n = \theta. \tag{1}$$

在这里，数 a_0 是不可能等于零的，否则，由(1)式便推出 $e_1, e_2, \cdots, e_n$ 线性相关，这是不可能的。因此，我们有

$$x = -\frac{a_1}{a_0}e_1 - \frac{a_2}{a_0}e_2 - \cdots - \frac{a_n}{a_0}e_n,$$

不妨将其写为

$$x = \xi_1 e_1 + \xi_2 e_2 + \cdots + \xi_n e_n. \tag{2}$$

即 x 可以表示成 $e_1, e_2, \cdots, e_n$ 的线性组合。如果 x 还有另一表示：

$$x = \eta_1 e_1 + \eta_2 e_2 + \cdots + \eta_n e_n, \tag{3}$$

则由(2)式减(3)式即得

$$(\xi_1 - \eta_1)e_1 + (\xi_2 - \eta_2)e_2 + \cdots + (\xi_n - \eta_n)e_n = \theta,$$

因基向量 $e_1, e_2, \cdots, e_n$ 线性无关，所以

$$\xi_1 - \eta_1 = 0, \quad \xi_2 - \eta_2 = 0, \cdots, \xi_n - \eta_n = 0,$$

从而有 $\xi_i = \eta_i \quad (i = 1, 2, \cdots, n)$，这证明了表示式的唯一性，定理证毕。

表示式(2)中的数 $\xi_1, \xi_2, \cdots, \xi_n$ 称为向量 x 在基 $e_1, e_2, \cdots, e_n$ 下的坐标。这定理说明，取定一组基后，每个向量 x 在这组基下的坐标是唯一确定的。x 的第 i 个坐标 $\xi_i \quad (i = 1, 2, \cdots, n)$ 也称为 x 的第 i 个分量。

我们再来看看前述几个例子中线性空间 P^n，$P^{m \times n}$ 及 $P[t]_n$ 的维数。

首先，容易证明

$$e_1 = (1,0,\cdots,0),\ e_2 = (0,1,0,\cdots,0),\cdots,$$
$$e_n = (0,0,\cdots,0,1)$$

是线性空间 P^n 的 n 个线性无关向量，又显然 P^n 中任一向量

$$x = (\xi_1,\xi_2,\cdots,\xi_n)$$

都可由这 n 个线性无关向量线性表出：

$$x = \xi_1 e_1 + \xi_2 e_2 + \cdots + \xi_n e_n,$$

因此，$x, e_1, e_2, \cdots, e_n$ 线性相关。从而得知 P^n 是 n 维线性空间。今后用得较多的是 R^n 及 C^n。

再考察线性空间 $P^{m\times n}$。若用 E_{ij} 表示第 i 行、第 j 列上的元素等于 1 而其他元素均等于零的 $m\times n$ 矩阵，则下列的 mn 个矩阵

$$E_{11}, E_{12}, \cdots, E_{ij}, \cdots, E_{mn}$$

$(i=1,2,\cdots,m;\ j=1,2,\cdots,n)$ 构成 $P^{m\times n}$ 的一组基，故 $P^{m\times n}$ 是 mn 维线性空间。今后用得较多的是 $R^{m\times n}$ 及 $C^{m\times n}$，包括它们当 $m=n$ 时的特殊情况。

最后，由于 $1, t, t^2, \cdots, t^{n-1}$ 是 $P[t]_n$ 的一组基，故 $P[t]_n$ 是 n 维线性空间。

§2 基变换与坐标变换

设 V 是数域 P 上的 n 维线性空间，又 $e_1, e_2, \cdots, e_n$ 及 $e'_1, e'_2, \cdots, e'_n$ 是 V 的两组基。假设这两组基的关系为

$$\begin{cases} e'_1 = a_{11}e_1 + a_{21}e_2 + \cdots + a_{n1}e_n \\ e'_2 = a_{12}e_1 + a_{22}e_2 + \cdots + a_{n2}e_n \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ e'_n = a_{1n}e_1 + a_{2n}e_2 + \cdots + a_{nn}e_n, \end{cases} \quad (1)$$

这组关系式可缩写为

$$
\left.\begin{aligned}
e'_i &= \sum_{k=1}^{n} a_{ki} e_k \\
i &= 1, 2, \cdots, n
\end{aligned}\right\} \qquad (1')
$$

现设 x 为 V 中任一向量，且在此二组基下其表示式为：

$$
x = \sum_{k=1}^{n} \xi_k e_k \qquad (2)
$$

$$
x = \sum_{i=1}^{n} \xi'_i e'_i. \qquad (3)
$$

下面我们要研究向量 x 在基变换下，其坐标的变化规律. 与变换(1)相对应，我们把矩阵

$$
A = \begin{bmatrix}
a_{11} & a_{12} & \cdots & a_{1n} \\
a_{21} & a_{22} & \cdots & a_{2n} \\
\cdots & \cdots & \cdots & \cdots \\
a_{n1} & a_{n2} & \cdots & a_{nn}
\end{bmatrix}
$$

称为从基 $e_1, e_2, \cdots, e_n$ 到基 $e'_1, e'_2, \cdots, e'_n$ 的过渡矩阵。由于基向量线性无关，并利用齐次线性方程组只有零解的条件，便可证明矩阵 A 是可逆的。由(3)式及(1′)式可得

$$
\begin{aligned}
x &= \sum_{i=1}^{n} \xi'_i e'_i \\
&= \sum_{i=1}^{n} \xi'_i \Big(\sum_{k=1}^{n} a_{ki} e_k \Big) \\
&= \sum_{k=1}^{n} \Big(\sum_{i=1}^{n} a_{ki} \xi'_i \Big) e_k,
\end{aligned} \qquad (4)
$$

由(2)式又有

$$
x = \sum_{k=1}^{n} \xi_k e_k. \qquad (5)
$$

由于 $e_1, e_2, \cdots, e_n$ 线性无关，所以(4)、(5)两式右边 e_k 的系数

应相等：

$$\xi_k = \sum_{i=1}^{n} a_{ki}\xi_i', \qquad k = 1,2,\cdots,n .$$

写成矩阵形式就是

$$\begin{bmatrix} \xi_1 \\ \xi_2 \\ \vdots \\ \xi_n \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix} \begin{bmatrix} \xi_1' \\ \xi_2' \\ \vdots \\ \xi_n' \end{bmatrix}$$

亦即

$$\begin{bmatrix} \xi_1 \\ \xi_2 \\ \vdots \\ \xi_n \end{bmatrix} = A \begin{bmatrix} \xi_1' \\ \xi_2' \\ \vdots \\ \xi_n' \end{bmatrix} \tag{6}$$

从而又有

$$\begin{bmatrix} \xi_1' \\ \xi_2' \\ \vdots \\ \xi_n' \end{bmatrix} = A^{-1} \begin{bmatrix} \xi_1 \\ \xi_2 \\ \vdots \\ \xi_n \end{bmatrix} \tag{7}$$

(6)、(7)两式给出了在基变换(1)下，向量 x 的坐标变换公式。

§3 子空间与维数定理

线性空间有些性质需用子空间的性质来表达，所以研究线性空间的子空间是必要的。

定义 1 设 V 是数域 P 上的线性空间，W 是 V 的一个非空子集。如果 W 对于线性空间 V 所定义的加法运算及数量乘法运算，也构成数域 P 上的线性空间，则称 W 为 V 的线性子空间，简称子

空间。

从线性空间的定义很容易找到上述非空子集为 V 的子空间的充要条件。这就是下述定理。

定理 1 设 W 是数域 P 上线性空间 V 的非空子集，则 W 是 V 的线性子空间的充要条件是：

(1) 若 $x,y\in W$，则 $x+y\in W$；

(2) 若 $x\in W,\lambda\in P$，则 $\lambda x\in W$

换言之，线性空间 V 的非空子集 W 是子空间的充要条件，是 W 关于 V 中定义的两个运算是“封闭”的。

证明 条件的必要性是显然的，因为当 W 为 V 的子空间时，由定义 1 即知条件(1)与(2)自然是满足的。反过来，若定理 1 的两个条件已满足，则可推出零向量 $\theta\in W$(取 $\lambda=0$ 并利用条件(2))；又当 $x\in W$ 时，取 $\lambda=-1$ 便可从条件(2)推出 $-x\in W$。至于线性空间定义中的其他运算“规则”，由于对 V 中所有元素都成立，当然对子集 W 的元素也能成立。所以，定理中的条件也是充分的。证毕。

在线性空间 V 中，由单个的零向量组成的子集 $\{\theta\}$ 是 V 的一个子空间，称为*零子空间*，而线性空间 V 本身也是 V 的一个子空间。这两个子空间称为*平凡子空间*。零空间的维数定义为零。

例 1 在 n 维线性空间 P^n 中，子集

$$W=\{x\mid Ax=\theta,x\in P^n\}$$

构成 P^n 的一个 $n-r$ 维子空间，r 是 $A\in P^{m\times n}$的秩。

例 2 设 $x_1,x_2,\cdots,x_m$ 是数域 P 上线性空间 V 的 m 个向量。则这组向量的所有形如

$$\lambda_1x_1+\lambda_2x_2+\cdots+\lambda_mx_m\quad(\lambda_i\in P)$$

的线性组合构成的集合非空，且对 V 中的加法及数量乘法皆封闭，故形成 V 的一个子空间，称为由这组向量生成的子空间，并记为 $L(x_1,x_2,\cdots,x_m)$。

例 2 提供了构作已知线性空间的子空间的方法．下面两个定理其实也给出了获得新的子空间的方法．

定理 2 设 V_1,V_2 是数域 P 上线性空间 V 的两个子空间，则它们的交 $W=V_1\cap V_2$ 也是 V 的子空间．

证明 由于每个子空间都包含零向量，所以零向量必定属于这两个子空间的交，即 W 不会是空集．

现任取 $x,y\in W$，则 $x,y\in V_i$，而 V_i 是子空间，所以 $x+y\in V_i(i=1,2)$，从而 $x+y\in W$．

又对任一 $\lambda\in P$ 及任一 $x\in W$，则又有

$$\lambda x\in V_i\quad(i=1,2),$$

从而 $\lambda x\in W$．定理证毕．

定理 3 设 V_1,V_2 是数域 P 上线性空间 V 的两个子空间，则它们的和

$$V_1+V_2=\{x+y\mid x\in V_1,y\in V_2\}$$

也是 V 的子空间．

证明 首先 V_1+V_2 不是空集，因为零向量属于 V_1 及 V_2，且 $\theta=\theta+\theta\in V_1+V_2$．其次，如果 u,v 是 V_1+V_2 中任二向量，且设

$$u=x_1+y_1,\quad v=x_2+y_2,$$

这里 $x_1,x_2\in V_1,\quad y_1,y_2\in V_2$。由于 V_1,V_2 是子空间，故

$$x_1+x_2\in V_1,\quad y_1+y_2\in V_2.$$

从而

$$u+v=(x_1+x_2)+(y_1+y_2)\in V_1+V_2.$$

同样地，对任一 $\lambda\in P$，则有

$$\lambda u=\lambda x_1+\lambda y_1\in V_1+V_2,$$

即 V_1+V_2 是 V 的子空间。证毕．

由于子空间的交与和都满足交换律及结合律，所以还可定义有限个子空间的交与和，并把上述两个定理推广到有限多个子空间的情形，兹不赘述．

例 3 $L(x_1,x_2,\cdots,x_s)+L(y_1,y_2,\cdots,y_t)$
$=L(x_1,x_2,\cdots,x_s,y_1,y_2,\cdots,y_t)$。

在定理 3 中，我们给出了两个子空间的和 V_1+V_2 的定义。在子空间的和中，有一个情形特别重要，这就是下面定义的子空间的直和。

定义 2 设 V_1,V_2 是线性空间 V 的两个子空间，如果这两个子空间的和 $W=V_1+V_2$ 具有性质：每个 $u\in W$，分解式

$$u=x+y \quad (\text{其中 } x\in V_1,\quad y\in V_2)$$

是唯一的，则称子空间 V_1 与 V_2 的和 $W=V_1+V_2$ 为**直和**，并记为 $W=V_1\oplus V_2$。

例 4 设有四维线性空间 R^4 的三个子空间：

$V_1=\{(a,\ b,\ 0,\ 0)\ |a,\ b\in R\}$；

$V_2=\{(0,\ 0,\ c,\ 0)\ |c\in R\}$；

$V_3=\{(0,\ d,\ e,\ 0)\ |d,\ e\in R\}$。

则 $T=V_1+V_3$ 不是直和，因为 T 中有向量 $(1,\ 1,\ 1,\ 0)$，分解式不唯一：

$$(1,1,1,0)=(1,2,0,0)+(0,-1,1,0),$$
$$(1,1,1,0)=(1,0,0,0)+(0,1,1,0).$$

但 $S=V_1+V_2$ 则是直和。因为当 $u\in S$，则有

$$u=(a,b,0,0)+(0,0,c,0)$$
$$=(a,b,c,0).$$

若 u 还有另一表示：

$$u=(a_1,b_1,0,0)+(0,0,c_1,0)$$
$$=(a_1,b_1,c_1,0),$$

显然，$a_1=a$，$b_1=b$，$c_1=c$。故 S 中每个向量的分解式唯一，从而 S 是直和。

下面讨论子空间的交与和的维数。线性空间 V 的维数常记为 $\dim(V)$ 或 $\dim V$。

定理 4(维数公式)　设 V 是数域 P 上的 n 维线性空间,V_1,V_2 是它的两个子空间,则有维数公式:

$$\dim V_1 + \dim V_2 = \dim(V_1 + V_2) + \dim(V_1 \cap V_2),$$

或写作:

$$\dim(V_1 + V_2) = \dim V_1 + \dim V_2 - \dim(V_1 \cap V_2).$$

证明　我们假设

$$\dim V_1 = r,\ \dim V_2 = s,\ \dim\ (V_1 + V_2)\ = k,$$

$$\dim\ (V_1 \cap V_2)\ = t,$$

现在 $V_1 \cap V_2$ 中选取一组基 $e_1,e_2,\cdots,e_t$,并扩充它*,使 r 个线性无关向量

$$e_1,e_2,\cdots,e_t,e'_{t+1},\cdots,e'_r$$

成为 V_1 的一组基;同样地,使 s 个线性无关向量

$$e_1,e_2,\cdots,e_t,\varepsilon_{t+1},\cdots,\varepsilon_s$$

成为 V_2 的一组基。

如能证明

$$e_1,e_2,\cdots,e_t,e'_{t+1},\cdots,e'_r,\ \varepsilon_{t+1},\cdots,\varepsilon_s \qquad (1)$$

为 V_1+V_2 的一组基,那就有

$$\dim(V_1 + V_2) = r + s - t,$$

从而定理便得证。

首先证明 V_1+V_2 中任一向量可由组(1)线性表出。为此,任取 $\boldsymbol{u} \in V_1+V_2$,则有 $\boldsymbol{x} \in V_1$ 及 $\boldsymbol{y} \in V_2$,使得

$$\begin{aligned} \boldsymbol{u} &= \boldsymbol{x} + \boldsymbol{y} \\ &= (\lambda_1 e_1 + \cdots + \lambda_t e_t + \mu_{t+1} e'_{t+1} + \cdots + \mu_r e'_r) \\ &\quad + (\lambda'_1 e_1 + \cdots + \lambda'_t e_t + \gamma_{t+1}\varepsilon_{t+1} + \cdots + \gamma_s \varepsilon_s) \end{aligned}$$

---

* 因 $r \geqslant t$,若 $r > t$,则 $e_1,e_2,\cdots,e_t$ 还不能构成 V_1 的一组基,这时可从 V_1 再选一非零向量 e'_{t+1},如果 $e_1,e_2,\cdots,e_t,e'_{t+1}$ 线性无关,就将 e'_{t+1} 添加到向量组 $e_1,e_2,\cdots,e_t$ 中去,否则就不要 e'_{t+1} 而另选一个。继续下去即可得 V_1 的基。

$$=(\lambda_1+\lambda_1')e_1+\cdots+(\lambda_t+\lambda_t')e_t+\mu_{t+1}e_{t+1}'$$
$$+\cdots+\mu_r e_r'+\gamma_{t+1}\varepsilon_{t+1}+\cdots+\gamma_s\varepsilon_s,$$

亦即 u 可由向量组(1)线性表出。

现在再来证明向量组(1)是线性无关的。事实上，如果有

$$\lambda_1 e_1+\cdots+\lambda_t e_t+\mu_{t+1}e_{t+1}'+\cdots+\mu_r e_r'+\gamma_{t+1}\varepsilon_{t+1}+$$
$$\cdots+\gamma_s\varepsilon_s=\theta,$$

则可得

$$\lambda_1 e_1+\cdots+\lambda_t e_t+\mu_{t+1}e_{t+1}'+\cdots+\mu_r e_r'$$
$$=-\gamma_{t+1}\varepsilon_{t+1}-\cdots-\gamma_s\varepsilon_s. \qquad (2)$$

这等式左边确定了一个属于 V_1 的向量 v，而由右边又可见 v 亦属于 V_2，从而 $v\in V_1\cap V_2$，故 v 可由 $V_1\cap V_2$ 的基 $e_1,e_2,\cdots,e_t$ 线性表出：

$$v=a_1e_1+a_2e_2+\cdots+a_te_t. \qquad (3)$$

故由(2)、(3)两式得：

$$a_1e_1+a_2e_2+\cdots+a_te_t+\gamma_{t+1}\varepsilon_{t+1}+\cdots+\gamma_s\varepsilon_s=\theta,$$

由此即得

$$a_1=\cdots=a_t=\gamma_{t+1}=\cdots=\gamma_s=0.$$

代入(3)式即得 $v=\theta$，再应用(2)式且因基向量线性无关，于是又得

$$\lambda_1=\cdots=\lambda_t=\mu_{t+1}=\cdots=\mu_r=0.$$

这已证明了向量组(1)线性无关。于是定理得证。

推论 若 n 维线性空间 V 的两个子空间的维数之和大于 n，则 $V_1\cap V_2$ 必含非零向量。

证明 由所设条件有 $\dim V_1+\dim V_2>n$，又 $\dim(V_1+V_2)\leqslant n$ 显然成立，故由维数公式即得

$$\dim(V_1\cap V_2)=\dim V_1+\dim V_2-\dim(V_1+V_2)>0,$$

所以 $V_1\cap V_2$ 含有非零向量。证毕。

若 $V_1\cap V_2=\{\theta\}$，则维数公式便成为

$$\dim(V_1+V_2)=\dim V_1+\dim V_2,$$

即和的维数等于维数的和。下述定理说明了 $V_1\cap V_2=\{\theta\}$ 与直和的关系。

定理 5 线性空间 V 的子空间 V_1 与 V_2 的和 V_1+V_2 是直和的充要条件是 $V_1\cap V_2=\{\theta\}$。

证明 若 $V_1\cap V_2$ 含有非零向量 $\boldsymbol{u}$，则 V_1+V_2 不是直和。这是因为由

$$\boldsymbol{u}+(-\boldsymbol{u})=\boldsymbol{\theta}=\boldsymbol{\theta}+\boldsymbol{\theta}\quad(\boldsymbol{u}\neq\boldsymbol{\theta})$$

推知零向量 $\boldsymbol{\theta}$ 有两种不同的分解方式。所以，若 V_1+V_2 是直和，则必定 $V_1\cap V_2=\{\theta\}$。

反之，如有 $V_1\cap V_2=\{\theta\}$，我们来证 V_1+V_2 是直和。事实上，如果 V_1+V_2 不是直和，则 V_1+V_2 中至少有一向量 $\boldsymbol{u}$ 有两种分解式：

$$\left.\begin{aligned}\boldsymbol{u}&=\boldsymbol{x}+\boldsymbol{y}\qquad(\boldsymbol{x}\in V_1,\quad \boldsymbol{y}\in V_2)\\ \boldsymbol{u}&=\boldsymbol{x}_1+\boldsymbol{y}_1\quad(\boldsymbol{x}_1\in V_1,\quad \boldsymbol{y}\in V_2)\end{aligned}\right\}\tag{1}$$

这里 $\boldsymbol{x}\neq\boldsymbol{x}_1$，$\boldsymbol{y}\neq\boldsymbol{y}_1$ 至少有一个成立。

设 $\boldsymbol{x}\neq\boldsymbol{x}_1$，则由(1)中两式相减即得

$$(\boldsymbol{x}-\boldsymbol{x}_1)+(\boldsymbol{y}-\boldsymbol{y}_1)=\boldsymbol{\theta},$$

即 $\boldsymbol{x}-\boldsymbol{x}_1=-(\boldsymbol{y}-\boldsymbol{y}_1)$；但是

$$\boldsymbol{x}-\boldsymbol{x}_1\in V_1,\ \boldsymbol{y}-\boldsymbol{y}_1\in V_2,\ -(\boldsymbol{y}-\boldsymbol{y}_1)\in V_2.$$

所以

$$\boldsymbol{\theta}\neq(\boldsymbol{x}-\boldsymbol{x}_1)\in V_1\cap V_2。$$

此与 $V_1\cap V_2=\{\theta\}$ 相矛盾，故 V_1+V_2 是直和。

推论 V_1+V_2 是直和的充要条件，是零向量的分解式唯一。

定理 6 若 V_1 与 V_2 是 n 维线性空间 V 的两个子空间，又 V_1+V_2 是直和，则

$$\dim(V_1+V_2)=\dim V_1+\dim V_2.$$

这是显然的。此等式亦可写成

$$\dim(V_1\oplus V_2)=\dim V_1+\dim V_2.$$

子空间直和概念可以推广到有限多个子空间的情形。而定理

6 的结果可以推广为

$$\dim(V_1 \oplus V_2 \oplus \cdots \oplus V_k) = \sum_{i=1}^{k} \dim V_i。$$

其证明从略。

§4 线性空间的同构

我们会遇到一些看起来不相同的线性空间，比如 P^n 及 $P[t]_n$，前者的一般元素为 n-数组：

$$x = (a_1, a_2, \cdots, a_n) \quad (a_i \in P)。$$

后者的一般元素为多项式：

$$p(t) = a_0 + a_1 t + \cdots + a_{n-1} t^{n-1} \quad (a_i \in P)。$$

那末这两个线性空间有无“本质”上的区别呢？

在线性空间的定义里，决定集合 V 能否构成数域 P 上一个线性空间，主要是看它是否定义了“加法”运算及“数量乘法”运算，以及这两个运算是否满足八条“规则”。而这两个运算的具体定义及集合 V 的元素具体是什么，在我们的讨论中是可以不考虑的。代数是关于运算规则的科学。具有相同运算规则的系统，在某种意义上就认为是相同的，可以不加区别的。确切地说，我们有下述定义。

定义 1 数域 P 上的两个线性空间 V 与 V' 称为是**同构的**，如果 V 与 V' 之间有一个一一对应 σ，使得对任何 x，$y \in V$ 及 $\lambda \in P$ 均满足：

(1) $\sigma(x+y) = \sigma(x) + \sigma(y)$；

(2) $\sigma(\lambda x) = \lambda \sigma(x)$。

σ 就称为从 V 到 V' 的**同构映射**。

我们也可以这样说，线性空间 V 与 V' 称为是同构的，如果二者之间能建立起元素（向量）的一一对应，并且这个对应保持 V 中

的加法运算及数量乘法运算，即在这个对应下，V 中向量 x，y 的和 $x+y$ 对应着 V' 中向量 x' 与 y' 的和 $x'+y'$；而且 V 中的数量乘积 λx 对应于 V' 中的数量乘积 $\lambda x'$（在这里我们设 V 中的 x，y 对应于 V' 中的 x'，y'），亦即，若 $x \to x'$，$y \to y'$，则 $x+y \to x'+y'$；$\lambda x \to \lambda x'$。

现在来讨论数域 P 上 n 维线性空间 V 与线性空间 P^n 的关系。在 V 中取定一组基 e_1，e_2，…，e_n；又 x，y 为 V 中任二向量，λ 为 P 中任意数，则有

$$x = \xi_1 e_1 + \xi_2 e_2 + \cdots + \xi_n e_n;$$

$$y = \eta_1 e_1 + \eta_2 e_2 + \cdots + \eta_n e_n;$$

$$\lambda x = (\lambda\xi_1) e_1 + (\lambda\xi_2) e_2 + \cdots + (\lambda\xi_n) e_n.$$

设 σ 为 V 到 P^n 的一个映射，它使任何 x，$y \in V$ 按下面方式同 P^n 中的元素对应起来：

$$x \to (\xi_1, \xi_2, \cdots, \xi_n);$$

$$y \to (\eta_1, \eta_2, \cdots, \eta_n).$$

则容易证明这个对应是 V 到 P^n 的一一对应，并且有

$$x + y \to (\xi_1, \xi_2, \cdots, \xi_n) + (\eta_1, \eta_2, \cdots, \eta_n);$$

$$\lambda x \to \lambda(\xi_1, \xi_2, \cdots, \xi_n).$$

因此，σ 是同构映射。于是，我们得到下述结论：

数域 P 上每个 n 维线性空间 V，当取定一组基后，V 与 P^n 同构。

由定义 1 看出，同构映射 σ：$V \to V'$ 具有下列基本性质：

1) $\sigma(\theta)=\theta$；　$\sigma(-\chi)=-\sigma(\chi)$；

2) $\sigma(\sum_{i=1}^{m} \lambda_i x_i) = \sum_{i=1}^{m} \lambda_i \sigma(x_i)$；

3) 若 x_1，x_2，…，x_m 为 V 中线性无关组，则 $\sigma(x_1)$，$\sigma(x_2)$，…，$\sigma(x_m)$ 在 V' 中线性无关；反之亦成立。即在同构对应下，线性无关组对应线性无关组。

证明 设有

$$\sum_{i=1}^{m}\lambda_i\sigma(x_i)=\theta,$$

则得

$$\sigma(\sum_{i=1}^{m}\lambda_i x_i)=\theta .$$

但由 1) 已知 $\sigma(\theta)=\theta$,而 σ 是个一一对应，故只能有

$$\sum_{i=1}^{m}\lambda_i x_i=\theta .$$

但 x_1，x_2，…，x_m 线性无关；所以得知

$$\lambda_1=\lambda_2=\cdots=\lambda_m=0 .$$

这就证明了向量组 $\sigma(x_1),\sigma(x_2),\cdots,\sigma(x_m)$线性无关。反之，若 x_1，x_2，…，x_m 是 V 中任意 m 个向量，但如果 $\sigma(x_1),\sigma(x_2),\cdots,\sigma(x_m)$ 是 V'中一线性无关组，则我们就可以证明原来的 m 个向量 x_1，x_2，…，x_m 在 V 中也一定是线性无关的。事实上，设有

$$\sum_{i=1}^{m}\lambda_i x_i=\theta, \qquad (*)$$

则由 $\sigma(\theta)=(\theta)$,及 2) 便可推知

$$\sum_{i=1}^{m}\lambda_i\sigma(x_i)=\theta,$$

但由假设，$\sigma(x_1)$，$\sigma(x_2)$，…，$\sigma(x_m)$线性无关，故由此式即得 $\lambda_1=\lambda_2=\cdots=\lambda_m=0$，联系到（*）式便知 x_1，x_2，…，x_m 线性无关。

4）同构的有限维线性空间，其维数相同。

证明 **因为有限维线性空间的维数就是它的最大线性无关组所含向量的个数。设 V 与 V'是两个同构的有限维线性空间，V 到 V'的同构映射为 σ。设 V 是 n 维的，x_1，x_2，…，x_n 是 V 的一个最大线性无关组。由性质 3) 知 $\sigma(x_1)$，$\sigma(x_2)$，…，$\sigma(x_n)$是 V'的 n 个线性无关向量。假如它还不是 V'的最大线性无关组，则把它扩充成最大线性无关组**

$$\sigma(x_1),\sigma(x_2),\cdots,\sigma(x_n),\sigma(x_{n+1}),\cdots,\sigma(x_{n+k}).$$

(因为 σ 是 V 到 V' 上的映射，故 V' 中的任一向量 x' 在 V 中都有原象 x，使得 $x'=\sigma(x)$。)而由性质 3)推出向量组

$$x_1,x_2,\cdots,x_n,\quad x_{n+1},\cdots,x_{n+k}$$

线性无关。这是一个矛盾，因此，

$$\sigma(x_1),\sigma(x_2),\cdots,\sigma(x_n)$$

是 V' 的最大线性无关组，故 V' 的维数是 n。

定理 1 数域 P 上的任意两个 n 维线性空间 V 与 V' 都是同构的。

证明 在 V 中选取一组基 $e_1,e_2,\cdots,e_n$，则 V 中任一向量 x，可表示为

$$x=\xi_1e_1+\xi_2e_2+\cdots+\xi_ne_n。\qquad(1)$$

又设 e_1', e_2', $\cdots$, e_n' 为 V' 的一组基，现作向量 x' 如下：

$$x'=\xi_1e_1'+\xi_2e_2'+\cdots+\xi_ne_n'.$$

则显然 $x'\in V'$。因此，V 中任一向量 x 都对应着 V' 中一个确定的向量 x'，且 x' 由 x 唯一确定。又由于两个空间在构成上完全平等，所以每个 $x'\in V'$，也能对应着 V 中一个唯一确定的向量 x。

由上面建立的对应法则易知，若 $x\to x'$，且 $y\to y'$，则

$$x+y\to x'+y';\quad \lambda x\to\lambda x'.$$

因此，V 与 V' 是同构的。

推论 数域 P 上两个有限维线性空间同构的充要条件，是维数相同。

我们小结一下这一节。由前面的说明及以上关于同构的讨论，可以知道，同构的线性空间有相同的代数性质（指那些仅与线性空间定义中两个运算有关的性质，而同构映射是保持这两个运算的），因此同构的线性空间是可以不加区别的，即认为是相同的。其次，由于一切 n 维线性空间（数域 P 上的）都与 P^n 同构，这就可以用较具体的 P^n 来认识较抽象的 n 维线性空间，并且 P^n 中许多

性质照样可以搬到一般 n 维线性空间里来。

§5 线性变换的概念

设 V 是数域 P 上的线性空间. 这里把从 V 到 V 的映射称为 V 的变换，线性变换是其中最简单、最基本的一种变换，它与矩阵、线性空间等都有密切联系，是矩阵论的主要研究对象之一。

如无特别指出，以下几节提到的线性空间，都是指数域 P 上的线性空间。

定义 1 线性空间 V 的一个变换 T 称为**线性变换**，如果对任意的 x，$y\in V$ 及 $\lambda\in P$，都有

$$T(x+y)=T(x)+T(y);$$

$$T(\lambda x)=\lambda T(x)。$$

由这定义可见，线性变换 T 是 $V\to V$ 的"保持向量加法"及"数量乘法"的变换。线性变换 T 的定义也可以叙述为：

若 T 是 V 到 V 的映射，如果在 T 作用下，$x\to x'$，$y\to y'$，则有

$$x+y\to x'+y',\ \lambda x\to \lambda x'\ .$$

这里 x，y 为 V 中任意元，λ 为 P 中任意元，则 T 便称为线性空间 V 的一个线性变换。

我们把上述的 $T(x)$或 x'称为向量 $x\in V$ 在线性变换 T 下的象，而 x 叫 $T(x)$或 x'的原象。我们约定：V 的两个线性变换 T 与 S 认为是相等的，当且只当对任何 $x\in V$ 均有 $T(x)=S(x)$。

我们先看几个线性变换的例子。

例 1 对每个 $x=(\xi_1,\ \xi_2,\ \xi_3,\ \xi_4)\in R^4$，由等式

$$T(x)=(\xi_1+\xi_2-3\xi_3-\xi_4,\quad 3\xi_1-\xi_2-3\xi_3+4\xi_4,0,0)$$

定义的变换 T 是 R^4 的线性变换。

例 2 设 B，C 是 $R^{n\times n}$的两个给定的矩阵，如果对任一 $X\in R^{n\times n}$，定义 T 为：

$$T(X) = BXC,$$

则 T 是线性空间 $R^{n\times n}$ 的线性变换。

例 3 在实多项式空间 $R[t]$ 中，由

$$T(p(t)) = \frac{\mathrm{d}}{\mathrm{d}t}p(t) \quad (p(t) \in R[t])$$

定义的变换(导数) T 是线性变换。

例 4 在由闭区间 $[a, b]$ 上全体连续函数构成的实线性空间中，由

$$T(f(t)) = \int_a^t f(u)\mathrm{d}u \quad (a < t \leqslant b)$$

定义的变换也是线性变换。

例 5 把线性空间 V 的每个向量都映射到零向量的变换叫做**零变换**；又把 V 中每个向量都映射到自身的变换叫做**单位变换**。易知这两个变换都是线性变换。

线性变换有下列简单性质：

(1) 若 T 是线性变换，则

$$T(\theta) = \theta; \qquad T(-x) = -T(x) .$$

这是因为：

$$T(\theta) = T(0x) = 0T(x) = \theta;$$

$$T(-x) = T[(-1)x] = (-1)T(x) = -T(x).$$

(2) 线性变换 T 保持向量的线性组合与线性关系式，即：

$$y = \sum_{i=1}^{m} \lambda_i x_i \Rightarrow T(y) = \sum_{i=1}^{m} \lambda_i T(x_i);$$

$$\sum_{i=1}^{m} \lambda_i x_i = \theta \Rightarrow \sum_{i=1}^{m} \lambda_i T(x_i) = \theta.$$

(3) 线性变换把线性相关向量组变成线性相关向量组.

这两个性质的证明是容易的，留给读者作为练习。注意，由(3)不能认为线性变换都能把线性无关组变为线性无关组。一个简单例子是零变换，它把任何线性无关组变成线性相关组 $\{\theta\}$。

现在我们来讨论线性变换的运算。设 V 是数域 P 上的线性空间，T_1，T_2，T_3 是 V 的三个线性变换，我们定义下列三种运算：

1）**线性变换的和**　对每个 $x\in V$，则满足

$$T(x)=T_1(x)+T_2(x)$$

的变换 T 称为线性变换 T_1 与 T_2 的和，并记作 $T=T_1+T_2$。

易证 T_1+T_2 也是 V 的线性变换。

2）**线性变换的乘积**　对每个 $x\in V$，满足

$$T(x)=T_1(T_2(x))$$

的变换 T 称为线性变换 T_1 与 T_2 的乘积，记为 $T=T_1T_2$，则 T 也是线性变换。

证明　对任何 x，$y\in V$ 及 $\lambda\in P$，我们有：

$$\begin{aligned}(T_1T_2)(x+y)&=T_1(T_2(x+y))\\&=T_1(T_2(x)+T_2(y))\\&=T_1(T_2(x))+T_1(T_2(y))\\&=(T_1T_2)(x)+(T_1T_2)(y);\\T_1T_2(\lambda x)&=T_1(T_2(\lambda x))\\&=T_1(\lambda T_2(x))\\&=(\lambda T_1)(T_2(x))=(\lambda T_1T_2)(x).\end{aligned}$$

3）**线性变换的数量乘法**　对每个 $x\in V$，$\lambda\in P$，满足

$$T(x)=\lambda(T_1(x))$$

的变换 T 称为数 λ 与线性变换 T_1 的数量乘积（法），记为 $T=\lambda T_1$。易证 λT_1 也是线性变换。又，$(-1)\ T_1$ 简记为 $-T_1$。

上述三种运算是线性变换的基本运算，这些运算具有下列性质：

（1）对线性空间 V 的任意三个线性变换 T_1，T_2，T_3，结合律成立：

$$T_1(T_2T_3)=(T_1T_2)T_3。$$

（2）V 中线性变换的加法满足交换律及结合律。

(3) V 中线性变换的乘法对加法的分配律成立。

(4) V 的零变换 O 及 V 的任一线性变换 T，满足关系式：

$$T+O=T,\quad T+(-T)=O\text{。}$$

(5) 数量乘法满足以下关系式：

$$(\lambda\mu)T=\lambda(\mu T),$$

$$(\lambda+\mu)T=\lambda T+\mu T,$$

$$\lambda(T_1+T_2)=\lambda T_1+\lambda T_2,$$

$$1T=T\text{。}$$

这里 λ，$\mu\in P$，T_1，T_2 及 T 为 V 的任意线性变换。

由于零变换及单位变换都是线性变换，所以线性空间 V 的所有线性变换组成的集合不会是空集，因而由以上的讨论得知：数域 P 上的线性空间 V 的全体线性变换组成的集合，对于线性变换的加法及数量乘法，也构成数域 P 上的一个线性空间，并用 $L(V)$ 来表示。

我们来研究线性变换的逆变换问题。

定义 2 设 I 为线性空间 V 的单位线性变换，T 为 V 的线性变换。如果存在 V 的一个线性变换 S，使得

$$TS=ST=I,$$

则称线性变换 T 是可逆的，而 S 称为 T 的逆变换，记为 T^{-1}。

读者可以证明：当线性变换 T 可逆时，其逆变换 T^{-1} 也是线性变换。当然，正如矩阵那样，并非每个线性变换都是可逆的。

例 6 设 T 是线性空间 V 的线性变换，则

$$T(V)=\{T\boldsymbol{v}\mid\boldsymbol{v}\in V\}$$

是 V 的子空间，称为象子空间。$T(V)$ 的维数叫做线性变换 T 的秩。

证明留给读者。

例 7 设 T 是线性空间 V 的线性变换，则集合

$$K=\{\boldsymbol{v}\in V\mid T\boldsymbol{v}=\boldsymbol{\theta}\}$$

是 V 的子空间，证明留给读者。这个子空间称为线性变换 T 的核(kernel)，并记为 $\ker(T)$ 或 $T^{-1}(\theta)$（这只是代表 θ 的原象组成的集合，而不表示 T 的可逆）。

由例 6 及例 7 可以进一步推得下述定理。

定理 1 设 T 是 n 维线性空间 V 的线性变换，则有维数关系：

$$\dim T(V) + \dim T^{-1}(\theta) = n .$$

证明 设 $\dim T^{-1}(\theta)=s$，又 $e_1, e_2, \cdots, e_s$ 是核 $T^{-1}(\theta)$ 的一组基。我们将它扩充，使

$$e_1, e_2, \cdots, e_s, \quad \varepsilon_1, \varepsilon_2, \cdots, \varepsilon_t$$

成为 V 的一组基，显然 $s+t=n$。如能证明

$$\dim T(V) = t,$$

则定理便得证。现设 x 是 V 的任一向量，则有

$$x = \sum_{i=1}^{s}\xi_i e_i + \sum_{j=1}^{t}\eta_j\varepsilon_j .$$

由于 $T(e_i)=\theta \quad (i=1, 2, \cdots, s)$，所以

$$T(x) = \sum_{j=1}^{t}\eta_j T(\varepsilon_j) .$$

当然，$T(x)\in T(V)$，所以上式表示 $T(V)$ 中的任一向量都是向量组

$$T(\varepsilon_1), T(\varepsilon_2), \cdots, T(\varepsilon_t) \tag{1}$$

的线性组合。现证向量组 (1) 线性无关。设有

$$\sum_{i=1}^{t}c_i T(\varepsilon_i) = \theta, \tag{2}$$

则有

$$T(\sum_{i=1}^{t}c_i\varepsilon_i) = \theta,$$

所以

$$y = \sum_{i=1}^{t}c_i\varepsilon_i \in T^{-1}(\theta),$$

从而 y 可由 T^{-1} (θ) 的基 e_1，e_2，…，e_s 线性表出，即

$$y = \sum_{j=1}^{s} d_j e_j ,$$

因此得：

$$\sum_{i=1}^{t} c_i \varepsilon_i - \sum_{j=1}^{s} d_j e_j = \theta \ . \qquad (3)$$

但已知 e_1，e_2，…，e_s，　ε_1，ε_2，…，ε_t 是 V 的一组基，故 (3) 式中一切 $c_i=0$，一切 $d_j=0$ ($i=1$，2，…，t；$j=1$，2，…，s)。因此，我们从 (2) 式导致了 $c_1=c_2=\cdots=c_t=0$。这证明了向量组 (1) 是线性无关的。

这样，我们证明了

$$T(\varepsilon_1),T(\varepsilon_2),\cdots,T(\varepsilon_t)$$

是 T (V) 的一组基，从而 T (V) 的维数等于 t。故定理得证。

§6　线性变换的矩阵表示

我们在上节已经得知：线性空间 V 的所有线性变换组成的集合，对于线性变换的加法及数量乘法，也构成一个线性空间 L (V)。若 V 是 n 维线性空间，那末 L (V) 的维数是多少？它与空间 $P^{n\times n}$ 有什么关系？这就是本节要讨论的主要问题。

我们先来证明一个定理。

定理 1　设 V 是数域 P 上的一个 n 维线性空间，e_1，e_2，…，e_n 是它的一组基，又 g_1，g_2，…，g_n 是 V 的任意 n 个向量。则存在唯一的一个线性变换 T，使得

$$T(e_1) = g_1,T(e_2) = g_2,\cdots,T(e_n) = g_n. \qquad (1)$$

证明　先证唯一性。若除了满足条件 (1) 的线性变换 T 外，还有线性变换 S 也满足同样的条件：

$$S(e_1) = g_1,S(e_2) = g_2,\cdots,S(e_n) = g_n.$$

现任取 $x \in V$，且设

$$x = \sum_{i=1}^{n} \xi_i e_i,$$

则有

$$T(x) = T(\sum_{i=1}^{n} \xi_i e_i) = \sum_{i=1}^{n} \xi_i T(e_i) = \sum_{i=1}^{n} \xi_i g_i;$$

同理有

$$S(x) = \sum_{i=1}^{n} \xi_i g_i.$$

即对任一 $x \in V$，都有 $T(x) = S(x)$，所以 $T = S$。这就证明了唯一性。

下面来证存在性．对 V 中任一向量 x：

$$x = \sum_{i=1}^{n} \xi_i e_i,$$

我们由等式

$$T(x) = \sum_{i=1}^{n} \xi_i g_i \tag{2}$$

定义 V 的一个变换 T。

T 显然是 V 的一个变换。易证 T 是线性变换．现取 $x = e_i$，则由（2）式即得

$$T(e_i) = g_i \quad (i = 1, 2, \cdots, n).$$

即线性变换 T 是满足条件（1）的。证毕。

此定理的唯一性部分说明：一个线性变换完全被它的一组基上的作用所决定．

为简单起见，以后我们用 Tx 代替 $T(x)$。

下面我们着手研究线性变换和矩阵的关系，亦即本节开头提出的主要问题。

设 V 是数域 P 上的 n 维线性空间，T 是 V 的一个线性变换。现取定 V 的一组基 $e_1, e_2, \cdots, e_n$，则每个 Te_i 都是 V 中向量（$i=1, 2, \cdots, n$），故可设：

$$\begin{cases} Te_1 = a_{11}e_1 + a_{21}e_2 + \cdots + a_{n1}e_n \\ Te_2 = a_{12}e_1 + a_{22}e_2 + \cdots + a_{n2}e_n \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ Te_n = a_{1n}e_1 + a_{2n}e_2 + \cdots + a_{nn}e_n. \end{cases} \tag{3}$$

我们把这 n 个向量等式缩写为如下形式：

$$Te_i = \sum_{k=1}^{n} a_{ki}e_k \tag{3'}$$

$$(i = 1,2,\cdots,n) \quad (a_{ki} \in P).$$

把矩阵

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix}$$

称为线性变换 T 在基 e_1，e_2，…，e_n 下的矩阵。

由此可见，在线性空间 V 取定一组基后，V 的每一个线性变换 T（$\in L(V)$）对应着一个矩阵 A（$\in P^{n\times n}$），其对应方式由（3）式反映出来。现把这个对应关系写为 $T\to A$，由定理 1 可推知这个对应是一一对应。并且，如果 $T_1\to A_1$，$T_2\to A_2$，则有

$$T_1 + T_2 \to A_1 + A_2, \quad \lambda T_1 \to \lambda A_1.$$

这由（3）式是很易推得的。这里 $\lambda\in P$ 是任意数。这样，我们已建立了下述定理。

定理 2 数域 P 上 n 维线性空间 V 的所有线性变换构成的线性空间 $L(V)$，在取定 V 的一组基之下，它与数域 P 上一切 $n\times n$ 矩阵所构成的线性空间 $P^{n\times n}$ 是同构的。

推论 $\dim L(V) = \dim P^{n\times n} = n^2$.

前面已经讲过，同构的线性空间有相同的代数性质，因而同构的空间可以看作是一样的。于是，$P^{n\times n}$ 的许多性质对于 $L(V)$ 也是成立的。不仅如此，应用线性变换与矩阵的对应关系，我们还

可以证明下述性质。

定理 3 设 $e_1, e_2, \cdots, e_n$ 是数域 P 上 n 维线性空间 V 的一组基，在这组基下，按照 (3) 式建立的线性变换同矩阵的对应关系，则有：

1）线性变换的乘积对应于矩阵的乘积；

2）可逆线性变换对应的矩阵也可逆，且逆变换对应于逆矩阵。

证明 1）设 T，S 是线性空间 V 的两个线性变换，在所取定的基下，它们对应的矩阵分别是 A，B。即是说，对于 $i=1, 2, \cdots, n$，有

$$Te_i = \sum_{k=1}^{n} a_{ki} e_k;$$

$$Se_i = \sum_{k=1}^{n} b_{ki} e_k.$$

因此得

$$\begin{aligned}(TS)e_i &= T(Se_i) = T\left(\sum_{k=1}^{n} b_{ki} e_k\right) \\ &= \sum_{k=1}^{n} b_{ki} T(e_k) \\ &= \sum_{k=1}^{n} b_{ki}\left(\sum_{t=1}^{n} a_{tk} e_t\right) \\ &= \sum_{t=1}^{n}\left(\sum_{k=1}^{n} a_{tk} b_{ki}\right) e_t \\ &= \sum_{t=1}^{n} c_{ti} e_t.\end{aligned}$$

这里记

$$c_{ti} = \sum_{k=1}^{n} a_{tk} b_{ki}, \quad (t, i = 1, 2, \cdots, n).$$

由矩阵乘积的定义，此式已表明 n 阶矩阵

$$C=(c_{ti})_{n\times n}=AB\ .$$

因此，线性变换的乘积 TS 在所取基下的矩阵是 AB。换言之，当 $T\to A$ 及 $S\to B$ 时，则有 $TS\to AB$。

2）又显然单位线性变换 I 在基 $e_1,e_2,\cdots,e_n$ 下的矩阵是单位矩阵 E，即变换 I 对应于矩阵 E。所以，当有 $ST=TS=I$ 时，便有 $BA=AB=E$。定理证毕。

最后，研究当基改变时，线性变换的矩阵的变化规律。

定理 4 设 V 是数域 P 上 n 维线性空间，$e_1,e_2,\cdots,e_n$ 及 $e'_1,e'_2,\cdots,e'_n$ 是 V 的两组基，从前一组基到后一组基的过渡矩阵是 C。又设 T 是 V 的一个线性变换，它在前后两组基下的矩阵分别是 A 与 B，则有 $B=C^{-1}AC$。

证明 由假设有

$$\begin{cases} e'_i=\sum\limits_{k=1}^{n}c_{ki}e_k, & i=1,2,\cdots,n \\ Te_i=\sum\limits_{k=1}^{n}a_{ki}e_k, & i=1,2,\cdots,n \\ Te'_i=\sum\limits_{k=1}^{n}b_{ki}e'_k, & i=1,2,\cdots,n. \end{cases}$$

因此得到

$$Te'_i=\sum_{k=1}^{n}b_{ki}\left(\sum_{t=1}^{n}c_{tk}e_t\right)=\sum_{t=1}^{n}\left(\sum_{k=1}^{n}c_{tk}b_{ki}\right)e_t,$$

以及

$$\begin{aligned} Te'_i&=T\left(\sum_{k=1}^{n}c_{ki}e_k\right)=\sum_{k=1}^{n}c_{ki}Te_k \\ &=\sum_{k=1}^{n}c_{ki}\sum_{t=1}^{n}a_{tk}e_t \\ &=\sum_{t=1}^{n}\left(\sum_{k=1}^{n}a_{tk}c_{ki}\right)e_t. \end{aligned}$$

比较这两式的右边，即有

$$\sum_{k=1}^{n} c_{tk} b_{ki} = \sum_{k=1}^{n} a_{tk} c_{ki}。$$

这等式左边正是矩阵 CB 的 (t, i) 元素，而右边也恰好是矩阵 AC 的 (t, i) 元素，对所有 $t, i=1, 2, \cdots, n$ 等式都成立。因此，有 $CB=AC$，即是 $B=C^{-1}AC$。证毕。

定义 1 若 $A, B \in P^{n\times n}$，如果存在可逆矩阵 $C \in P^{n\times n}$，使得

$$B = C^{-1}AC,$$

则称矩阵 A 相似于矩阵 B，并记作 $A \sim B$。这时也简单地说 A 与 B 相似。

由定义易知，矩阵的相似是等价关系，即相似具有下述三个性质：

1）自反性　$A \sim A$；

2）对称性　若 $A \sim B$，则 $B \sim A$；

3）传递性　若 $A \sim B$ 且 $B \sim C$，则 $A \sim C$。

以上 $A, B, C \in P^{n\times n}$ 皆为任意矩阵，读者可自证之。

于是定理 4 表明一个重要事实：一个线性变换在不同基下的矩阵是相似的。反过来也可以证明，两个相似矩阵总可以看成某一线性变换在两组不同基下的矩阵。

相似矩阵的概念及一些简单性质，读者已经熟知，兹不赘述。

线性变换和矩阵的上述相互关系是很重要的。线性空间、线性变换及矩阵三者之间都有着密切联系，熟习这种联系，对深入研究矩阵理论是很有益处的。

§7　不变子空间

定义 1 设 T 是线性空间 V 的一个线性变换，又 W 是 V 的一个子空间。若对任一 $x \in W$，都有 $Tx \in W$，亦即

$$T(W) \subseteq W,$$

则称W为T的不变子空间，也说成是子空间W对线性变换T是不变的。

由定义可知，零空间及V本身都是T的不变子空间。

现设V_1，V_2是n维线性空间V的两个子空间，且都是线性变换T的不变子空间．如果

$$V = V_1 \oplus V_2,$$

且e_1，e_2，…，e_m与e_{m+1}，…，e_n分别是V_1与V_2的一组基，则向量组

$$e_1, e_2, \cdots, e_m, \quad e_{m+1}, \cdots, e_n \tag{1}$$

便构成V的一组基．由于V_1，V_2对T不变，所以有

$$\begin{aligned}
Te_1 &= a_{11}e_1 + \cdots + a_{m1}e_m \\
&\cdots\cdots \\
Te_m &= a_{1m}e_1 + \cdots + a_{mm}e_m \\
Te_{m+1} &= a_{m+1,m+1}e_{m+1} + \cdots + a_{n,m+1}e_n \\
&\cdots\cdots\cdots \\
Te_n &= a_{m+1,n}e_{m+1} + \cdots + a_{nn}e_n
\end{aligned}$$

因此线性变换T在基（1）下的矩阵为分块对角矩阵

$$A = \begin{pmatrix} A_1 & \\ & A_2 \end{pmatrix}$$

这里

$$A_1 = \begin{pmatrix} a_{11} & \cdots & a_{1m} \\ \cdots & \cdots & \cdots \\ a_{m1} & \cdots & a_{mm} \end{pmatrix}, \quad A_2 = \begin{pmatrix} a_{m+1,m+1} & \cdots & a_{m+1,n} \\ \cdots & \cdots & \cdots \\ a_{n,m+1} & \cdots & a_{nn} \end{pmatrix}$$

易知，若$V = V_1 \oplus V_2 \oplus \cdots \oplus V_k$，又$T$为$V$的线性变换，且每个$V_i$都对$T$不变，则适当选择基，变换$T$在此基下的矩阵便为分块对角形：

$$A=\begin{bmatrix} A_1 & & & \\ & A_2 & & \\ & & \ddots & \\ & & & A_k \end{bmatrix} \tag{2}$$

还可以证明，若 V 可分解为 k 个子空间 V_i（$i=1, 2, \cdots, k$）的直和，则存在 V 的一个线性变换 T，使每个 V_i 都是 T 的不变子空间，从而 T 在某组基下的矩阵具有分块对角形（2）的形式。

显然，若 n 维线性空间 V 可分解为线性变换 T 的 n 个一维不变子空间的直和，则 T 的矩阵可以具有对角形矩阵的形状。对角线上的元素就是线性变换 T 所对应的矩阵 A 的特征值，亦称为线性变换 T 的特征值。

本节的讨论说明线性空间分解为直和、线性变换以及分块对角形矩阵的关系．再一次说明线性空间、线性变换及矩阵三者是息息相关的。

习　题　一

1．在 n 维线性空间 P^n 中，下列 n 维向量的集合 V，是否构成 P 上的线性空间：

(1) $V=\{(a, b, a, b, \cdots, a, b) \mid a, b\in P\}$；

(2) $V=\{(a_1, a_2, \cdots, a_n) \mid \sum_{i=1}^{n} a_i=1\}$；

(3) $V=\{x=(\xi_1, \xi_2, \cdots, \xi_n)^T \mid Ax=\theta, A\in P^{n\times n}\}$。

2．按通常矩阵的加法及数与矩阵的乘法，下列的数域 P 上方阵集合是否构成 P 上线性空间：

(1) 全体形如 $\begin{pmatrix} 0 & a \\ -a & b \end{pmatrix}$ 的二阶方阵的集合；

(2) 全体 n 阶对称(或反对称，上三角)矩阵的集合；

(3) $V=\{X \mid AX=\theta, X\in P^{n\times n}\}$　（A 为给定的 n 阶方阵，$A\in P^{n\times n}$）。

3．证明：线性空间定义中，第 3)个条件的第(1)个不是独立

的，亦即它可由其余七个“规则”推出。

4. 证明：对于有限维线性空间，在线性空间定义中的“规则” $1x=x$ 是可以证明的，而不必在定义中给出。

5. 证明：若线性空间 V 中的每个向量都可由 V 中 n 个向量 $e_1,e_2,\cdots,e_n$ 线性表出，且有一个向量表示法唯一，则 V 必是 n 维空间，且这组向量是它的一组基。

6. 在三维线性空间 P^3 中，分别求下面的向量 α 在基 $\varepsilon_1,\varepsilon_2,\varepsilon_3$ 下的坐标：

1) $\alpha=(1,2,1)$,

$\varepsilon_1=(1,1,1),\varepsilon_2=(1,1,-1),\varepsilon_3=(1,-1,-1)$.

2) $\alpha=(3,7,1)$,

$\varepsilon_1=(1,3,5),\varepsilon_2=(6,3,2),\varepsilon_3=(3,1,0)$.

7. 在 R^4 中有两组基：

1) $\alpha_1=(1,0,0,0)$, $\alpha_2=(0,1,0,0)$,

$\alpha_3=(0,0,1,0)$, $\alpha_4=(0,0,0,1)$;

2) $\beta_1=(2,1,-1,1),\beta_2=(0,3,1,0)$,

$\beta_3=(5,3,2,1)$, $\beta_4=(6,6,1,3)$.

试求(1) 从第1)组基到第2)组基的过渡矩阵；

(2) 向量 $x=(\xi_1,\xi_2,\xi_3,\xi_4)$ 对第2)组基的坐标；

(3) 对两组基有相同坐标的非零向量。

8. 在 R^n 中，分量满足下列条件的全体向量能否构成 R^n 的子空间：

(1) $\xi_1+\xi_2+\cdots+\xi_n=0$;

(2) $\xi_1+\xi_2+\cdots+\xi_n=2$.

9. 试证：在 R^4 中，由 $(1,1,0,0),(1,0,1,1)$ 生成的子空间与由 $(2,-1,3,3),(0,1,-1,-1)$ 生成的子空间相同。

10. 设 V_1,V_2 都是线性空间 V 的子空间，且 $V_1\subseteq V_2$，证明：如果 $\dim V_1=\dim V_2$，则 $V_1=V_2$。

11. 求 R^4 的子空间

$V=\{(a_1,a_2,a_3,a_4)\mid a_1-a_2+a_3-a_4=0\}$,

$W=\{(a_1,a_2,a_3,a_4)\mid a_1+a_2+a_3+a_4=0\}$

的交 $V\cap W$ 的一组基。

12. 设向量组

$$\alpha_1=(1,0,2,1),\alpha_2=(2,0,1,-1),\alpha_3=(3,0,3,0),$$

$$\beta_1=(1,1,0,1),\beta_2=(4,1,3,1)。$$

若 $V_1=L(\alpha_1,\alpha_2,\alpha_3)$,$V_2=L(\beta_1,\beta_2)$,求 V_1+V_2 的维数及一组基。

13. 设 $e_1,e_2,\cdots,e_n$ 及 $\varepsilon_1,\varepsilon_2,\cdots,\varepsilon_n$ 是 n 维线性空间 V 的两组基,证明:

(1) 在两组基上坐标完全相同的全体向量的集合 V_1 是 V 的子空间;

(2) 若空间 V 的每个向量在这两组基上的坐标完全相同,则 $e_i=\varepsilon_i,i=1,2,\cdots,n$。

14. 设 V_1,V_2 分别是齐次线性方程组 $x_1+x_2+\cdots+x_n=0$ 与 $x_1=x_2=\cdots=x_n$ 的解空间,试证明 $P^n=V_1\oplus V_2$。

15. 设 A 是任一 $m\times n$ 矩阵,将 A 任意分块成

$$A=\begin{bmatrix}A_1\\A_2\\\vdots\\A_s\end{bmatrix}$$

证明:n 元齐次线性方程组 $Ax=\theta$ 的解空间 V 是齐次线性方程组 $A_ix=\theta$ 的解空间 $V_i(i=1,2,\cdots,s)$的交 $V=V_1\cap V_2\cap\cdots\cap V_s$。

16. 证明:每个 n 维线性空间都可以表示成 n 个一维子空间的直和。

17. 证明:$T_1(x_1,x_2)=(x_2,-x_1)$,$T_2(x_1,x_2)=(x_1,-x_2)$是 R^2 的两个线性变换,并求 T_1+T_2、T_1T_2 及 T_2T_1。

18. 对任一 $A\in P^{n\times n}$,又给定 $C\in P^{n\times n}$,定义变换 T 如下:

$$T(A)=CA-AC,$$

证明:

1) T 是 $P^{n\times n}$的线性变换;

2) 对任意 $A,B\in P^{n\times n}$有

$$T(AB)=T(A)\cdot B+A\cdot T(B)。$$

19. 设 T,S 是 R^3 的二个线性变换,它们定义为:

$$T(x,y,z)=(x+y+z,0,0),S(x,y,z)=(y,z,x).$$

试证 $T+S$ 的象集是 R^3,即$(T+S)(R^3)=R^3$。

20. 设 T 是 R^3 的线性变换,它定义为

$$T(x,y,z)=(0,x,y),$$

求 T^2 的象集及核。

21. 在 R^3 中，求下列各线性变换 T 在所指定基下的矩阵：

1) $T(x_1,x_2,x_3)=(2x_1-x_2,x_2+x_3,x_1)$在基

$$\varepsilon_1=(1,0,0),\varepsilon_2=(0,1,0),\varepsilon_3=(0,0,1)$$

下的矩阵；

.2) 已知线性变换 T 在基

$$\eta_1=(-1,1,1),\eta_2=(1,0,-1),\eta_3=(0,1,1)$$

下的矩阵为

$$\begin{bmatrix} 1 & 0 & 1 \\ 1 & 1 & 0 \\ -1 & 2 & 1 \end{bmatrix},$$

求 T 在基 $\varepsilon_1=(1,0,0),\varepsilon_2=(0,1,0),\varepsilon_3=(0,0,1)$下的矩阵。

22. 给定线性空间 R^3 的两组基：

$$\varepsilon_1=(1,0,1),\varepsilon_2=(2,1,0),\varepsilon_3=(1,1,1),$$

$$\eta_1=(1,2,-1),\eta_2=(2,2,-1),\eta_3=(2,-1,-1).$$

又设 T 是 R^3 的线性变换，且 $T\varepsilon_i=\eta_i\quad(i=1,2,3)$。试求：

(1) 从基$\{\varepsilon_i\}$到基$\{\eta_i\}$(三元集)的过渡矩阵；

(2) T 在基$\{\varepsilon_i\}$下的矩阵；

(3) T 在基$\{\eta_i\}$下的矩阵。

23. 若矩阵 A 可逆，证明 AB 与 BA 相似。

24. 若 $A\sim B,C\sim D$，试证

$$\begin{pmatrix} A & \\ & C \end{pmatrix}\sim\begin{pmatrix} B & \\ & D \end{pmatrix}.$$

25. 1) 证明 $T(x_1,x_2,\cdots,x_n)=(0,x_1,\cdots,x_{n-1})$是线性空间 P^n 的线性变换，且 $T^n=0$(零变换)

2) 求 T 的核 $T^{-1}(\theta)$的维数及象集 $T(V)$的维数。

第二章　内积空间

上章所讲的线性空间，其基本运算是向量的加法及向量与数域中的数的数量乘法，并无考虑过向量的长度、向量之间的夹角等度量性质。由于各种需要，我们有必要在线性空间中引入内积概念，使得在这样的空间里可以处理这些度量性质的问题，深化对线性空间、线性变换的研究。

§1　内积空间的概念

定义 1　设 V 是实数域 R 上的线性空间。如果对 V 中任意两个向量 $\boldsymbol{x},\boldsymbol{y}$ 都有一个实数（记为 (x,y)）与它们相对应，并且满足下列各个条件，则实数 (x,y) 称为向量 $\boldsymbol{x},\boldsymbol{y}$ 的内积：

(1) $(x,y)=(y,x)$；

(2) $(\lambda x,y)=\lambda(x,y)$，　$(\lambda\in R)$

(3) $(x+y,z)=(x,z)+(y,z)$，　$(z\in V)$

(4) $(x,x)\geqslant 0$，当且仅当 $\boldsymbol{x}=\theta$ 时，等号成立。

而线性空间 V 则称为实内积空间，简称内积空间。

平常空间向量的内积（亦称数量积）显然适合定义 1 中列举的所有性质，故空间向量的全体关于向量加法及数与向量的乘法，就构成一个内积空间。

例 1　若对 n 维线性空间 R^n 中的任二向量

$$\boldsymbol{x}=(\xi_1,\xi_2,\cdots,\xi_n),\boldsymbol{y}=(\eta_1,\eta_2,\cdots,\eta_n)$$

定义内积为

$$(x,y)=\sum_{i=1}^{n}\xi_i\eta_i,$$

则容易验证它满足内积定义的条件，从而 R^n 成为一个内积空间，我们仍用 R^n 来表示它。

内积空间 R^n 称为欧几里得(Euclid)空间，简称欧氏空间。稍后我们就知道，所有 n 维（实）内积空间都与 R^n 同构，因此我们把有限维（实）内积空间也都称为欧氏空间。本书所讲的空间主要是指有限维空间，虽然无论在线性空间的定义中，或在实内积空间的定义中，都没有规定空间的维数是有限的，甚至在讨论一些简单性质时，也无此规定，但是在本书或一般的线性代数及矩阵论的书中，主要的、绝大部分的内容都是对有限维空间来说的。

例 2 考虑 n^2 维线性空间 $R^{n\times n}$。如果对任何 $A,B\in R^{n\times n}$，定义

$$(A,B)=\sum_{i,j=1}^{n}a_{ij}b_{ij},$$

则容易验证(A,B)满足内积定义的各个条件，从而 $R^{n\times n}$构成一内积空间。

由定义可推得内积(x,y)具有下列基本性质：

(1) $(x,\lambda y)=\lambda(x,y)$；

(2) $(x,y+z)=(x,y)+(x,z)$；

(3) $(x,\theta)=(\theta,y)=0$。

其中(3)的证明如下：

$$(x,\theta)=(x,0y)=0(x,y)=0\ .$$

有了内积概念，我们就可以在内积空间中引入向量的长度及向量之间的夹角等概念，我们先来证明关于内积的一个重要不等式。

定理 1 设 V 是内积空间，x,y 是 V 中任二向量，则有

$$(x,y)^2\leqslant(x,x)(y,y),$$

且等号当且只当 x,y 线性相关时成立。

证明 设 t 为一任意实数，则由内积定义的条件(4)，得知向量内积

$$(x-ty,x-ty)\geqslant 0 .$$

也就是说，对任意实数 t，

$$(y,y)t^2-2(x,y)t+(x,x)\geqslant 0 .$$

此不等式左边是个关于 t 的二次三项式，对任意实数 t，它都取非负值，故其判别式

$$\Delta=[-2(x,y)]^2-4(y,y)(x,x)\leqslant 0 .$$

由此便得：

$$(x,y)^2\leqslant (x,x)(y,y).$$

如果 x,y 线性相关，不妨设 $y=\lambda x$（λ 是实数）. 此时，我们有

$$\begin{aligned}(x,y)^2&=(x,\lambda x)^2=[\lambda(x,x)]^2\\&=(x,x)(\lambda x,\lambda x)\\&=(x,x)(y,y) .\end{aligned}$$

同样地，若 $x=\mu y$，也可以证明

$$(x,y)^2=(x,x)(y,y).$$

因此，当 x,y 线性相关时，定理中的不等式成为等式. 反过来，如果等式 $(x,y)^2=(x,x)(y,y)$ 成立，则 x,y 必定线性相关. 事实上，若 x,y 线性无关，则对任何实数 t，都有 $x-ty\neq\theta$，从而就有

$$(x-ty,x-ty)>0 .$$

由定理前半部分的证明中可以看到，此时，前面提到的判别式 $\Delta<0$，从而导致

$$(x,y)^2<(x,x)(y,y) 。$$

这与 $(x,y)^2=(x,x)(y,y)$ 相矛盾，因而定理得证.

定理 1 中的不等式，常称为柯西-许瓦兹(Cauchy-Schwarz)不等式. 下面用到它时，我们简单地称之为(C. -S.)不等式.

现在我们来定义向量的长度.

定义 2 设 x 是内积空间 V 的任一向量，则非负实数 $\sqrt{(x,x)}$

称为向量 x 的长度，并记为 $|x|$，亦即我们定义向量 x 的长度为：

$$|x| = \sqrt{(x,x)} .$$

由于向量的内积 (x,y) 在内积空间中是个实数，因此利用长度概念，(C.—S.)不等式又可以表示为：

$$|(x,y)| \leqslant |x| \cdot |y| .$$

当 x,y 都不是零向量时，由此不等式可得

$$\frac{|(x,y)|}{|x| \cdot |y|} \leqslant 1,$$

亦即

$$-1 \leqslant \frac{(x,y)}{|x| \cdot |y|} \leqslant 1 .$$

因此，我们可以用等式

$$\cos\varphi = \frac{(x,y)}{|x| \cdot |y|}$$

来定义两个非零向量 x,y 的夹角 φ，且限制 φ 的取值范围为 $0 \leqslant \varphi \leqslant \pi$。当 $(x,y)=0$ 时，则称 x,y 是正交的*，且记为 $x \perp y$。

例 3 若 x,y 是两个正交向量，则有

$$|x+y|^2 = |x|^2 + |y|^2.$$

一般地，如果 $x_1,x_2,\cdots,x_k$ 是 k 个两两正交的向量，则有

$$|x_1 + x_2 + \cdots + x_k|^2 = |x_1|^2 + |x_2|^2 + \cdots + |x_k|^2.$$

这利用内积性质及正交条件是不难证明的，故留给读者作为练习。

从定理 1 可以推出如下简单推论。

推论 对内积空间 V 的任二向量 x,y 都有：

(1) $|x+y| \leqslant |x| + |y|$；

(2) $|x-y| \geqslant |x| - |y|$。

---

* 由于 $(x,\theta)=0$，所以认为零向量与任何向量是正交的。故由 $(x,y)=0$ 定义 $x \perp y$ 时，不要求 $x,y \neq \theta$。

证明 因为

$$
\begin{aligned}
|x+y|^2 &= (x+y, x+y) \\
&= (x,x) + 2(x,y) + (y,y) \\
&\leqslant (x,x) + 2|x|\cdot|y| + (y,y) \\
&= |x|^2 + 2|x|\cdot|y| + |y|^2 \\
&= (|x| + |y|)^2,
\end{aligned}
$$

由此即得$|x+y|\leqslant|x|+|y|$。又应用这一结果得到等式$x=(x-y)+y$,即得

$$|x| = |(x-y)+y| \leqslant |x-y| + |y|,$$

因此得$|x|-|y|\leqslant|x-y|$。这就是(2)式。

把定理 1 应用到欧氏空间 R^n 即得不等式:

$$\left|\sum_{i=1}^{n}\xi_i\eta_i\right| \leqslant \sqrt{\sum_{i=1}^{n}\xi_i^2}\cdot\sqrt{\sum_{i=1}^{n}\eta_i^2}\ .$$

§2 正交基及子空间的正交关系

由于在内积空间中有向量夹角的概念,因而在各种基中,我们可以选择一种特殊的、使用起来很方便的基,这就是正交基。这种情况在普通的线性空间里是没有的。"正交"概念的引入,使得在内积空间中增加了不少在一般线性空间中所没有的、很有意义的性质。

内积空间中两两正交的一组非零向量,称为**正交组**。正交组是线性无关的。事实上,若 $x_1,x_2,\cdots,x_m$ 为一正交组,如果有

$$\lambda_1x_1 + \lambda_2x_2 + \cdots + \lambda_mx_m = \theta,$$

则用 x_i 与此向量等式两边作内积,注意到当 $i\neq j$ 时均有$(x_i,x_j)=0$,我们便可得到

$$\lambda_i(x_i, x_i) = (x_i, \theta) = 0 .$$

因 $x_i \neq \theta$,故$(x_i, x_i) > 0$,因此有 $\lambda_i = 0(i=1,2,\cdots,m)$,即 $x_1, x_2, \cdots, x_m$ 线性无关.

定义 1 在 n 维欧氏空间中,由正交组构成的基称为*正交基*;如果正交基中每个基向量的长度都等于单位长度,则这组正交基便称为*标准正交基*.

简单地说,若 $e_1, e_2, \cdots, e_n$ 是 n 维欧氏空间 V 的一组非零向量,且满足条件

$$(e_i, e_j) = \begin{cases} 1, & \text{当 } i = j; \\ 0, & \text{当 } i \neq j. \end{cases}$$

$(i, j = 1, 2, \cdots, n)$,则 $e_1, e_2, \cdots, e_n$ 即为一标准正交基.

定理 1 任一 n 维欧氏空间 V 都存在正交基.

证明 设 $f_1, f_2, \cdots, f_n$ 为 V 的一组基,我们从这组基出发,去构作 V 的一组正交基. 下面是具体作法.

首先可以取 $e_1 = f_1$,接着作向量

$$e_2 = f_2 + \alpha_1 e_1,$$

其中系数 α_1 可由正交条件$(e_2, e_1) = 0$ 来确定,由于

$$(e_2, e_1) = (f_2 + \alpha_1 e_1, e_1) = (f_2, e_1) + \alpha_1(e_1, e_1) = 0,$$

因此可得

$$\alpha_1 = -\frac{(f_2, e_1)}{(e_1, e_1)}。$$

因 $e_1 = f_1$ 与 f_2 线性无关,所以 $e_2 \neq \theta$,且以上求得的 e_1, e_2 是正交的. 类似地,令

$$e_3 = f_3 + \beta_1 e_1 + \beta_2 e_2,$$

则 e_3 必不为零向量. 再由正交条件(要求):

$$(e_3, e_1) = 0, \qquad (e_3, e_2) = 0,$$

便可求得系数 β_1, β_2 为:

$$\beta_1 = -\frac{(f_3, e_1)}{(e_1, e_1)}, \quad \beta_2 = -\frac{(f_3, e_2)}{(e_2, e_2)}.$$

因而 e_3 便可确定，于是又得到正交组 e_1, e_2, e_3. 按此方法做下去，若已作出正交组：

$$e_1, e_2, \cdots, e_{n-1}.$$

则令

$$e_n = f_n + \lambda_1 e_1 + \lambda_2 e_2 + \cdots + \lambda_{n-1} e_{n-1}.$$

显然 $e_n \neq \theta$. 再由正交条件

$$(e_n, e_1) = 0, \ (e_n, e_2) = 0, \ \cdots, \ (e_n, e_{n-1}) = 0,$$

便可定出各个 λ_i:

$$\lambda_i = -\frac{(f_n, e_i)}{(e_i, e_i)}, \quad i = 1, 2, \cdots, n-1 .$$

因此，我们便得到正交基 $e_1, e_2, \cdots, e_n$. 定理证毕.

注意，如果我们将上面求得的正交基的每个基向量都化为单位向量，即令

$$e_i' = \frac{e_i}{|e_i|}, \quad i = 1, 2, \cdots, n,$$

则得到标准正交基：$e_1', e_2', \cdots, e_n'$. 因此，我们从定理 1 又可得知，每个有限维内积空间都存在标准正交基. 定理 1 及上述作法，不仅解决了正交基的存在性问题，而且还提供了一个实际作法，这个作法叫做施密特(Schmidt)正交化过程*.

我们假设 V 是个 n 维欧氏空间，不妨设 $e_1, e_2, \cdots, e_n$ 是它的一组标准正交基. 现考察 V 中任二向量 x, y 在这组基下内积的表达式. 为此，可设

$$x = \xi_1 e_1 + \xi_2 e_2 + \cdots + \xi_n e_n,$$

$$y = \eta_1 e_1 + \eta_2 e_2 + \cdots + \eta_n e_n.$$

---

* 这个"过程"通常指到作出一标准正交基为止. 由于主要困难是作出正交基，即"正交化"这一步，故为提法方便，我们用此讲法.

于是，利用内积性质（稍作推广）及标准正交基的定义，可得：

$$(x,y)=(\xi_1e_1+\cdots+\xi_ne_n,\eta_1e_1+\cdots+\eta_ne_n)$$
$$=\xi_1\eta_1+\xi_2\eta_2+\cdots+\xi_n\eta_n。$$

由此可见，在标准正交基下，欧氏空间（有限维实内积空间）向量内积可由坐标的一个简单表达式来描述。

下面我们讨论一下：从一组标准正交基到另一组标准正交基的过渡矩阵有何特性。

设 $e_1,e_2,\cdots,e_n$ 及 $e_1',e_2',\cdots,e_n'$ 是 n 维欧氏空间 V 的两组标准正交基，从前一组基到后一组基的过渡矩阵为 A。即

$$e_i'=\sum_{k=1}^{n}a_{ki}e_k,\quad i=1,2,\cdots,n。$$

由于

$$(e_i',e_j')=(\sum_{k=1}^{n}a_{ki}e_k,\sum_{l=1}^{n}a_{lj}e_l)$$
$$=\sum_{k=1}^{n}\sum_{l=1}^{n}a_{ki}a_{lj}(e_k,e_l)$$
$$=\sum_{k=1}^{n}a_{ki}a_{kj},$$

另方面，由于

$$(e_i',e_j')=\begin{cases}1, & 当\ i=j\\0, & 当\ i\neq j\end{cases}$$

所以有

$$\sum_{k=1}^{n}a_{ki}a_{kj}=\begin{cases}1, & 当\ i=j\\0, & 当\ i\neq j\end{cases}$$
$$(i,j=1,2,\cdots,n)$$

这组关系式表明矩阵 A 满足等式：

$$A^TA=E。\qquad(1)$$

其中 A^T 是 A 的转置矩阵。读者已经知道，满足(1)式的 n 阶实矩阵 A 称为正交矩阵。因而，在 n 维欧氏空间中，从一组标准正交基到

另一组标准正交基的过渡矩阵是正交矩阵。

在这里，我们复述一下正交矩阵 A 的基本性质。由(1)式两边取行列式可得 $|A|=\pm 1$，故正交矩阵 A 是可逆的，且 $A^{-1}=A^T$，又易证 A^{-1} 也是正交矩阵，而且两个 n 阶正交矩阵的乘积还是正交矩阵。由正交矩阵 A 的定义又可以看到：A 的每个列(行)向量(看作欧氏空间 R^n 中的向量，下同)都是单位向量，且不同的两个列(行)向量是正交的。

注意，由于内积空间是线性空间，因而上章中关于子空间的有关结论，自然在这里仍能成立。下面我们只讨论一下那些与度量性质有关的子空间性质。

定义 2 设 V_1,V_2 是内积空间 V 的两个子空间。如果对任意的 $x\in V_1,y\in V_2$，都有

$$(x,y)=0,$$

则称 V_1 与 V_2 *是正交的*，并记为 $V_1\perp V_2$。

特别，如果 V 中某个向量 x，与子空间 V_1 的每个向量都正交，则称 x 与 V_1 正交，记为 $x\perp V_1$。

定理 2 内积空间 V 的两个子空间 V_1 与 V_2 如果是正交的，则它们的和 V_1+V_2 是直和。

证明 这只须证明 V_1+V_2 的零向量 θ 表示方式唯一。由于 $\theta=\theta+\theta$ 是一种表示，假如还有 $x\in V_1,y\in V_2$，使得 $x+y=\theta$，则有

$$0=(\theta,x)=(x+y,x)=(x,x)+(y,x)=(x,x),$$
$$(\text{因}(y,x)=0)\ .$$

于是有 $x=\theta$。同理可证 $y=\theta$。故零向量的表示方式唯一。证毕。

对有限多个子空间的情形，也有类似的结果。

定义 3 设 V_1,V_2 是内积空间 V 的两个子空间，且满足条件

$$V_1\perp V_2,\quad V_1+V_2=V,$$

则称 V_2 是 V_1 的*正交补空间*，简称*正交补*。

由定义易知，当 V_2 是 V_1 的正交补时，则 V_1 也是 V_2 的正交补。

定理 3 n 维欧氏空间 V 的任一子空间 V_1 都有唯一的正交补。

证明 若 $V_1=\{\theta\}$，则 V 就是 V_1 的正交补，并且唯一。现假设 $V_1\neq\{\theta\}$，因 V_1 在 V 的内积定义下也是个欧氏空间，故在 V_1 中可选取一组正交基 $e_1,e_2,\cdots,e_m$。这里 m 是 V_1 的维数。再从 V 中选取 $n-m$ 个向量添加进去，使

$$e_1,e_2,\cdots,e_m,\quad e_{m+1},\cdots,e_n$$

成为 V 的一组正交基。

不难证明，$L(e_{m+1},e_{m+2},\cdots,e_n)$ 就是 V_1 的正交补，将其记为 V_2。

再证唯一性。设除 V_2 外，还有 V_3 也是 V_1 的正交补。由正交补的定义且应用定理 2，即得

$$V=V_1\oplus V_2,$$
$$V=V_1\oplus V_3.$$

令 $x\in V_2$，则 $x\in V$，故由第二式有

$$x=x_1+x_3,$$

这里 $x_1\in V_1,x_3\in V_3$。又因 $x\perp x_1$，所以

$$0=(x,x_1)=(x_1+x_3,x_1)$$
$$=(x_1,x_1)+(x_3,x_1),$$

但 $x_3\perp x_1$，故 $(x_3,x_1)=0$，从而有 $(x_1,x_1)=0$，于是 $x_1=\theta$。由此得知 $x=x_3\in V_3$，即有 $V_2\subseteq V_3$。

同理可证 $V_3\subseteq V_2$。因此，有 $V_3=V_2$，故唯一性得证。证毕。

V_1 的正交补记为 $V_1^{\perp}$，于是由定理 3 有

$$\dim V_1+\dim V_1^{\perp}=n,$$

且 $V_1^{\perp}$ 恰由 V 中所有与 V_1 正交的向量所组成。

§3 内积空间的同构

与线性空间的同构类似，我们来讨论内积空间的同构问题。由于内积空间是定义了内积的线性空间，故下述定义是很自然的。

定义 1 两个内积空间 V 与 V' 称为是同构的，如果二者之间存在一个一一对应 σ，并且对任何 $x,y\in V$，$\lambda\in R$，下列条件都满足：

1) $\sigma(x+y)=\sigma(x)+\sigma(y)$；

2) $\sigma(\lambda x)=\lambda\sigma(x)$；

3) $(\sigma(x),\sigma(y))=(x,y)$。

这就是说，两个内积空间认为是同构的，首先作为线性空间它们是同构的，其次，在这个同构之下，向量内积是保持不变的。

由定义可知，同构的两个欧氏空间有相同的维数。

定理 1 所有 n 维欧氏空间都同构。

证明 设 V 是 n 维欧氏空间，又 $e_1,e_2,\cdots,e_n$ 是它的一组标准正交基，则任一 $x\in V$ 可表示为

$$x=\xi_1e_1+\xi_2e_2+\cdots+\xi_ne_n.$$

现由

$$\sigma(x)=(\xi_1,\xi_2,\cdots,\xi_n)\in R^n$$

定义一个 V 到 R^n 的映射 σ，则易知 σ 是个一一对应，且可以验证定义 1 中的条件 1)及 2)都是满足的，下面来证条件 3)亦满足。

再任取 V 中另一向量 $y=\eta_1e_1+\eta_2e_2+\cdots+\eta_ne_n$，则由 σ 的定义又有

$$\sigma(y)=(\eta_1,\eta_2,\cdots,\eta_n)\in R^n.$$

由 R^n 中向量内积定义，则有

$$(\sigma(x),\sigma(y))=\xi_1\eta_1+\xi_2\eta_2+\cdots+\xi_n\eta_n,$$

而 V 中向量 x,y 的内积 (x,y) 在这组标准正交基下的表示式为

$$(x,y)=\xi_1\eta_1+\xi_2\eta_2+\cdots+\xi_n\eta_n.$$

因此，有$(\sigma(x),\sigma(y))=(x,y)$。

综上所述，即知 V 与 R^n 同构。又不难证明欧氏空间的同构是等价关系*。因此所有 n 维欧氏空间都同构。证毕。

从这个定理可见，最基本、最简单的 n 维欧氏空间就是 R^n。

§4 正交变换

在内积空间中有一种特别的线性变换（正交变换），在许多场合都很有用，这种变换在欧氏空间（有限维实内积空间）中常可叙述为几种相互等价的提法。这些提法的每一种都刻画了正交变换的基本性质。

定义 1 设 T 是内积空间 V 的线性变换，若 T 能保持 V 中向量内积不变，即对任何 $x,y\in V$，都有

$$(Tx,Ty)=(x,y),$$

则线性变换 T 称为 V 的一个正交变换。

定理 1 设 T 是 n 维欧氏空间 V 的一个线性变换，则下列各个命题彼此等价：

(1) T 是正交变换；

(2) T 保持向量的长度不变，即对任一 $x\in V$，都有 $|Tx|=|x|$；

(3) 若 $e_1,e_2,\cdots,e_n$ 是 V 的一组标准正交基，则 $Te_1,Te_2,\cdots,Te_n$ 也是 V 的一组标准正交基；

(4) T 在 V 的任一标准正交基下的矩阵是正交矩阵。

证明 (1)⇔(2)：

若 T 是正交变换，则有

* 参考第一章§6。

$$(Tx,Tx) = (x,x),$$

两边开平方即得$|Tx|=|x|$。

反之,若 T 保持向量的长度不变,即

$$|Tx| = \sqrt{(Tx,Tx)} = |x| = \sqrt{(x,x)},$$

$$|Ty| = \sqrt{(Ty,Ty)} = |y| = \sqrt{(y,y)}.$$

因此,

$$\begin{cases}(Tx,Tx) = (x,x), \\ (Ty,Ty) = (y,y).\end{cases} \tag{*}$$

同样地,亦有

$$(T(x+y),T(x+y)) = (x+y,x+y).$$

把这式两边展开即得

$$(Tx,Tx) + 2(Tx,Ty) + (Ty,Ty)$$

$$=(x,x) + 2(x,y) + (y,y).$$

利用(*)式,便得

$$(Tx,Ty) = (x,y),$$

即 T 是正交变换。

(1)$\Leftrightarrow$(3):

设 T 是正交变换,则对 V 的任一组标准正交基 $e_1,e_2,\cdots,e_n$,都有

$$(Te_i,Te_j) = (e_i,e_j) = \begin{cases}1, & 当\ i=j \\ 0, & 当\ i\neq j\end{cases}$$

这里 $i,j=1,2,\cdots,n$。因此,$Te_1,Te_2,\cdots,Te_n$ 也是一组标准正交基。

反之,如果当 $e_1,e_2,\cdots,e_n$ 是标准正交基时,$Te_1,Te_2,\cdots,Te_n$ 也是标准正交基,则对 V 中任二向量

$$x = \xi_1 e_1 + \xi_2 e_2 + \cdots + \xi_n e_n,$$

$$y = \eta_1 e_1 + \eta_2 e_2 + \cdots + \eta_n e_n$$

便有

$$Tx = \xi_1 Te_1 + \xi_2 Te_2 + \cdots + \xi_n Te_n,$$

$$Ty = \eta_1 Te_1 + \eta_2 Te_2 + \cdots + \eta_n Te_n.$$

因此

$$(Tx, Ty) = \xi_1\eta_1 + \xi_2\eta_2 + \cdots + \xi_n\eta_n = (x, y),$$

从而 T 是正交变换.

(3)⇔(4):

设 T 在标准正交基 $e_1, e_2, \cdots, e_n$ 下的矩阵为 A,即是说,

$$Te_i = \sum_{k=1}^{n} a_{ki} e_k, \qquad i = 1, 2, \cdots, n$$

如果 $Te_1, Te_2, \cdots, Te_n$ 是标准正交基,则 A 可看成由标准正交基 e_1, $e_2, \cdots, e_n$ 到标准正交基 $Te_1, Te_2, \cdots, Te_n$ 的过渡矩阵,从而 A 是正交矩阵(见§2)。反之,若 A 是正交矩阵,则

$$\begin{aligned}(Te_i, Te_j) &= \left(\sum_{k=1}^{n} a_{ki} e_k, \sum_{l=1}^{n} a_{lj} e_l\right) \\ &= \sum_{k=1}^{n} a_{ki} a_{kj} \\ &= \begin{cases} 1, & \text{当 } i = j \\ 0, & \text{当 } i \neq j \end{cases}\end{aligned}$$

最后一步是因为 $A^T A = E$ (A 为正交矩阵),这里 $i, j = 1, 2, \cdots, n$. 这证明 $Te_1, Te_2, \cdots, Te_n$ 是标准正交基。定理证毕.

例 1 设 T 是欧氏空间 R^3 的线性变换,

$$T(\xi_1, \xi_2, \xi_3) = (\xi_2, \xi_3, \xi_1)$$

对任一$(\xi_1, \xi_2, \xi_3) \in R^3$ 成立,试证明 T 是正交变换.

证明 设 $x = (\xi_1, \xi_2, \xi_3) \in R^3$, 由定理 1, 我们只须证明 $|Tx| = |x|$. 由于

$$\begin{aligned}(Tx, Tx) &= ((\xi_2, \xi_3, \xi_1), (\xi_2, \xi_3, \xi_1)) \\ &= \xi_2^2 + \xi_3^2 + \xi_1^2 \\ &= (x, x),\end{aligned}$$

由此即得 $|Tx| = |x|$。

例2 1) 设 T 是内积空间 V 的一个线性变换。证明：T 是正交变换的充要条件是：T 保持任意两向量 x,y 的距离不变，即

$$|Tx - Ty| = |x - y|。$$

2) 问：内积空间的保持距离不变的变换是否一定是线性变换？

证明 1) 设 T 是正交变换，则 T 保持向量的长度不变。从而

$$|Tx - Ty| = |T(x - y)| = |x - y|。$$

反之，设对任意向量 x,y 有

$$|Tx - Ty| = |x - y|,$$

则取 $y=\theta$ 便有 $|Tx|=|x|$。即 T 保持向量长度不变，故 T 为正交变换。

2) 不一定。设 x_0 为 V 中某一固定非零向量，又令

$$Tx = x + x_0, \quad (对任意 x \in V)$$

则 T 是 V 的一个变换，并保持任二向量的距离不变：

$$|Tx - Ty| = |(x + x_0) - (y + x_0)| = |x - y|。$$

但是，T 显然不是线性变换（?）。（从而也不是正交变换。）

例3 设 T 是内积空间 V 的一个变换。证明：如果 T 保持向量的内积不变，即

$$(Tx,Ty) = (x,y), \quad x,y \in V$$

则 T 一定是线性变换，因而是正交变换。

证明 先证：$T(x+y)=Tx+Ty$。由于

$$\begin{aligned}
&(T(x + y) - Tx - Ty, T(x + y) - Tx - Ty) \\
=&(T(x + y),T(x + y)) - 2(T(x + y),Tx) - 2(T(x + y),Ty) \\
&+ (Tx,Tx) + (Ty,Ty) + 2(Tx,Ty) \\
=&(x + y,x + y) - 2(x + y,x) - 2(x + y,y) \\
&+ (x,x) + (y,y) + 2(x,y) \\
=&0
\end{aligned}$$

因此

$$T(x+y)-Tx-Ty=\theta,$$

从而有 $T(x+y)=Tx+Ty$。

再证：$T(\lambda x)=\lambda Tx$，（λ 为实数）

同上述方法类似，读者可验证等式

$$(T(\lambda x)-\lambda Tx, T(\lambda x)-\lambda Tx)=0。$$

从而有

$$T(\lambda x)-\lambda Tx=\theta,$$

此即 $T(\lambda x)=\lambda Tx$。

故 T 为线性变换，从而是正交变换。

§5 点到子空间的距离与最小二乘法

定义 1 设 V 是欧氏空间，又 x，$y\in V$，则向量 $x-y$ 的长度 $|x-y|$ 称为向量 x 与 y 的距离，并记为 $d(x,y)$。

不难验证向量距离的三个基本性质（对任何 $x,y,z\in V$）：

(1) $d(x,y)=d(y,x)$；

(2) $d(x,z)\leqslant d(x,y)+d(y,z)$；

(3) $d(x,y)\geqslant 0$，仅当 $x=y$ 时等号成立。

在初等几何里，我们知道点到直线（或平面）上所有点的距离以垂线最短。我们现在来证明：欧氏空间 V 的一个指定向量和一个子空间 W 的各个向量间的距离也以"垂线最短"。

设 $W=L(x_1,x_2,\cdots,x_k)$，又 $x\in V$ 为一指定向量。首先易知 $x\perp W\Leftrightarrow x\perp x_i\quad(i=1,2,\cdots,k)$。设 $y\in W$ 且满足条件 $(x-y)\perp W$，则对任一 $z\in W$，都有

$$|x-y|\leqslant|x-z|。$$

即是说，向量 x 到 W 的各个向量间的距离以"垂线"$|x-y|$ 最短。下面来证明这不等式。

因 $x-z=(x-y)+(y-z)$，又 $y-z\in W$，故 $(x-y)\perp(y-z)$。

因此，有（见§1，例3）：

$$|x-z|^2 = |x-y|^2 + |y-z|^2,$$

所以有$|x-y| \leqslant |x-z|$。

这个简单的几何事实可用来解决一些实际问题。其中的一个应用是解决最小二乘法问题。最小二乘法在系统理论中处理最小优化问题时有重要的应用。下面我们只是用欧氏空间的概念来表达最小二乘法，并给出最小二乘解所满足的代数条件。

设已给不相容线性方程组

$$Ax = b .$$

这里 $A=(a_{ij})_{n\times s}$，$b=(b_1,b_2,\cdots,b_n)^T$，$x=(x_1,x_2,\cdots,x_s)^T$（每个 x_i 都是实变数）。因为这方程组无解，我们只好设法找出一组数 $x_1^0,x_2^0,\cdots,x_s^0$，使平方偏差

$$\delta = \sum_{i=1}^{n}(a_{i1}x_1 + a_{i2}x_2 + \cdots + a_{is}x_s - b_i)^2$$

最小。这组数称为此方程组的最小二乘解，这一方法叫做最小二乘法。

令 $y=Ax$，则 y 当然是个 n 维列向量。上述平方偏差 δ 也就是 $|y-b|^2$，而最小二乘法就是要找一组数 $x_1^0,x_2^0,\cdots,x_s^0$，使 y 与 b 的距离最小。

设 $A=(\alpha_1,\alpha_2,\cdots,\alpha_s)$，$\alpha_i$ 表示 A 的第 i 列。则有

$$y = x_1\alpha_1 + x_2\alpha_2 + \cdots + x_s\alpha_s. \qquad (1)$$

显然，$y\in L(\alpha_1,\alpha_2,\cdots,\alpha_s)$，故最小二乘法可叙述为：

求 x 使 $|y-b|^2$ 最小，即在 $L(\alpha_1,\alpha_2,\cdots,\alpha_s)$ 中找一向量 y，使得向量 b 到它的距离比到子空间 $L(\alpha_1,\alpha_2,\cdots,\alpha_s)$ 中其他向量的距离都短。

由本节开头所讲的结论，若(1)式的 y 为所求的向量，则向量

$$c = b - y = b - Ax$$

必须垂直于子空间 $L(\alpha_1,\alpha_2,\cdots,\alpha_s)$，为此必须而且只须

$$(c,\alpha_1) = (c,\alpha_2) = \cdots = (c,\alpha_s) = 0 .$$

这条件相当于

$$\alpha_1^T c = 0,\ \alpha_2^T c = 0, \cdots,\ \alpha_s^T c = 0 .$$

这组等式相当于

$$A^T(\boldsymbol{b} - A\boldsymbol{x}) = \boldsymbol{\theta},$$

亦即

$$A^T A\boldsymbol{x} = A^T\boldsymbol{b} .$$

这就是最小二乘解所满足的代数方程。它是一个线性方程组，系数矩阵为 A^TA，常数项是列向量 $A^T\boldsymbol{b}$。

例 用最小二乘法解方程组：

$$\begin{cases} x_1 + x_2 = 1 \\ x_1 + x_3 = 2 \\ x_1 + x_2 + x_3 = 0 \\ x_1 + 2x_2 - x_3 = -1 \end{cases}$$

由于

$$A = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 1 & 1 \\ 1 & 2 & -1 \end{bmatrix},\ A^T = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 0 & 1 & 2 \\ 0 & 1 & 1 & -1 \end{bmatrix},\ \boldsymbol{b} = \begin{bmatrix} 1 \\ 2 \\ 0 \\ -1 \end{bmatrix}$$

所以

$$A^T A\boldsymbol{x} = \begin{bmatrix} 4 & 4 & 1 \\ 4 & 6 & -1 \\ 1 & -1 & 3 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 2 \\ -1 \\ 3 \end{bmatrix} = A^T\boldsymbol{b} .$$

于是求得最小二乘解为

$$x_1 = \frac{17}{6},\quad x_2 = -\frac{13}{6},\quad x_3 = -\frac{4}{6} .$$

§6 复内积空间(酉空间)

以上几节讲的是内积定义为一个实数的实内积空间，当它是有限维空间时，我们也把它叫作欧氏空间。虽然实内积空间的理论已在许多问题上扮演着重要角色，但是这还不足以解决所有的问题，复内积空间的引入才能弥补这一不足。在矩阵分析的应用中，复内积空间用途是很广的。

复内积空间是实内积空间的推广，许多概念、结论及证明方法都与以上所讲的相类似，因而在这一节我们不打算作详细的讨论，不必作类似性的重复。

定义 1 设 V 是复数域 C 上的线性空间，如果对 V 中任意两个向量 x,y 都有一个复数(记作 (x,y))与它们对应，且满足下列各个条件，则复数 (x,y) 称为 x,y 的内积：

(1) $(x,y)=\overline{(y,x)}$；

(2) $(\lambda x,y)=\lambda(x,y)$；

(3) $(x+y,z)=(x,z)+(y,z)$；

(4) $(x,x)\geqslant 0$，仅当 $x=\theta$ 时等号成立。

而 V 就称为复内积空间，或酉空间。这里的 x,y,z 是 V 中任意向量，λ 是任意复数。又 $\overline{(y,x)}$ 表示 (y,x) 的共轭复数。

注意，内积定义中的条件(1)保证 (x,x) 是实数。又若条件(1)不作这样规定，而仍旧用实内积空间定义中条件(1)的规定，就可能会出现 $x\neq\theta$ 但 $(x,x)=0$ 的情形，比如在下面的例子中，取 $x=(3,4,5i)$ 又 $\xi_1\eta_1+\xi_2\eta_2+\cdots+\xi_n\eta_n$，就会发生这种情形。

例 1 在 n 维线性空间 C^n 中，如果对 C^n 中任二向量 $x=(\xi_1,\xi_2,\cdots,\xi_n)$，$y=(\eta_1,\eta_2,\cdots,\eta_n)$，定义

$$(x,y)=\xi_1\bar{\eta}_1+\xi_2\bar{\eta}_2+\cdots+\xi_n\bar{\eta}_n.$$

则容易验证 (x,y) 满足内积定义各条件，从而 C^n 构成一酉空间。

注意酉空间 C^n 的上述内积定义又可简写为：

$$(x,y)=xy^H。$$

这里 x,y 均为行向量，y^H 表示 y 的共轭转置。又当 x,y 为列向量时，则有 $(x,y)=y^Hx$。这种写法以后是经常使用的。

在这个例子中，若定义 $(x,y)=\bar{\xi}_1\eta_1+\bar{\xi}_2\eta_2+\cdots+\bar{\xi}_n\eta_n$，则内积定义的条件(2)便不满足。

在在酉空间中，内积具有下列基本性质：

(1) $(x,\lambda y)=\bar{\lambda}(x,y)$;

(2) $(x,y+z)=(x,y)+(x,z)$;

(3) $(x,\theta)=(\theta,y)=0$。

这些性质与实内积空间的情形类似，证法也类似，故不重复。

在酉空间中，向量 x 的长度也定义为：

$$|x|=\sqrt{(x,x)}\ .$$

虽然，当 $(x,y)=0$ 时也称向量 x,y 为正交的，但在酉空间中不再定义向量间的夹角，因为向量的内积一般是复数。又若 $x,y\in C^n$（见例 1），则有

$$|x|=\sqrt{|\xi_1|^2+\cdots+|\xi_n|^2},|y|=\sqrt{|\eta_1|^2+\cdots+|\eta_n|^2}\ .$$

在实内积空间中，我们曾得到(C.-S.)不等式 $|(x,y)|\leqslant|x|\cdot|y|$。这不等式在酉空间中仍旧成立，而且等号也是当且只当 x,y 线性相关时成立。我们来证明这个事实。设 x,y 为酉空间 V 中任二向量，λ 为复数，则

$$(x-\lambda y,x-\lambda y)\geqslant 0\ , \qquad (1)$$

展开即为

$$(x,x)-\bar{\lambda}(x,y)-\lambda\overline{(x,y)}+\lambda\bar{\lambda}(y,y)\geqslant 0\ 。\qquad (2)$$

若 $y=\theta$，则我们想要建立的不等式

$$|(x,y)|\leqslant|x|\cdot|y|$$

显然适合，因为它的两边都是零。故设 $y\neq\theta$。现用数 $\dfrac{(x,y)}{(y,y)}$ 来代替不等式(2)中的 λ，且用正数 (y,y) 来乘不等式的各项，便可得出

$$(x,x)(y,y)-\overline{(x,y)}(x,y)-(x,y)\overline{(x,y)}+(x,y)\overline{(x,y)}\geqslant 0,$$

或

$$(x,y)\overline{(x,y)}\leqslant(x,x)(y,y),$$

即
$$|(x,y)|\leqslant|x|\cdot|y|. \qquad (3)$$

这就是酉空间中的(C.-S.)不等式。如果 x,y 线性无关，则 $x-\lambda y\neq\theta$，因而(1)式为严格不等式

$$(x-\lambda y,x-\lambda y)>0,$$

从而(3)式也是严格不等式

$$|(x,y)|<|x|\cdot|y|.$$

如果 x,y 线性相关，例如，$x=\lambda y$，则

$$|(x,y)|=|(\lambda y,y)|=|\lambda|(y,y)=|x|\cdot|y|.$$

这已完成了我们的证明。

对于例 1 的酉空间 C^n 的内积，则(C.-S.)不等式为

$$|\xi_1\bar{\eta}_1+\cdots+\xi_n\bar{\eta}_n|\leqslant\sqrt{|\xi_1|^2+\cdots+|\xi_n|^2}\cdot\sqrt{|\eta_1|^2+\cdots+|\eta_n|^2}.$$

应用上述(C.-S.)不等式，我们又可以证明：在酉空间中，不等式

$$|x+y|\leqslant|x|+|y|$$

也成立。

事实上，我们有

$$\begin{aligned}|x+y|^2&=(x+y,x+y)\\&=(x,x)+(x,y)+\overline{(x,y)}+(y,y)\\&=(x,x)+2\mathrm{Re}(x,y)+(y,y),\end{aligned}$$

其中 $\mathrm{Re}(x,y)$ 为 (x,y) 的实部，因

$$\mathrm{Re}(x,y)\leqslant|(x,y)|\leqslant|x|\cdot|y|,$$

所以

$$|x+y|^2\leqslant|x|^2+2|x|\cdot|y|+|y|^2=(|x|+|y|)^2,$$

于是有 $|x+y|\leqslant|x|+|y|$。

在酉空间中也可以定义正交基和标准正交基；正交组线性无关，从正交组出发去构作 n 维酉空间的一组正交基的正交化过程，在酉空间中仍然成立。

容易证明，在 n 维酉空间 V 中，在标准正交基 $e_1, e_2, \cdots, e_n$ 下任二向量

$$x = \xi_1 e_1 + \xi_2 e_2 + \cdots + \xi_n e_n,$$
$$y = \eta_1 e_1 + \eta_2 e_2 + \cdots + \eta_n e_n$$

的内积 (x, y) 可表示成：

$$(x, y) = \xi_1 \bar{\eta}_1 + \xi_2 \bar{\eta}_2 + \cdots + \xi_n \bar{\eta}_n.$$

将正交变换的概念推广到酉空间，便有

定义 2 若 T 为酉空间 V 的线性变换，且对任何 $x, y \in V$，都有

$$(Tx, Ty) = (x, y),$$

则 T 称为 V 的**酉变换**。即，酉空间的酉变换，是保持任二向量内积不变的线性变换。

同正交矩阵（实矩阵）相类似，对复数矩阵亦有相应的概念：

定义 3 若 $A \in C^{n \times n}$，且 $A^H A = AA^H = E$，则 A 称为**酉矩阵**。这里 A^H 是 A 的共轭转置。

当 A 为实矩阵时，酉矩阵 A 也就是正交矩阵。

同实内积空间的情形一样，我们可以证明下述定理。

定理 1 设 T 是 n 维酉空间 V 的线性变换，则下列各命题互相等价：

(1) T 是酉变换；

(2) T 保持向量的长度不变，即 $x \in V$ 时，都有 $|Tx| = |x|$；

(3) 若 $e_1, e_2, \cdots, e_n$ 是 V 的标准正交基，则 $Te_1, Te_2, \cdots, Te_n$ 也是 V 的标准正交基；

(4) T 在任一标准正交基下的矩阵是酉矩阵。

其证明与 § 4 定理 1 相仿，兹不赘述。

酉矩阵 A 具有下列基本性质：

1) A 的行列式的模等于 1；

2) $A^{-1}=A^H,(A^{-1})^H=A=(A^H)^{-1}$；

3) A^{-1}也是酉矩阵，两个 n 阶酉矩阵的乘积也是酉矩阵；

4) A 的每个列(行)向量(看作酉空间 C^n 的向量，下同)是单位向量；不同的两个列(行)向量是酉正交的(在 C^n 的内积定义下正交)。

同实内积空间情况类似，复内积空间也有同构的概念，而且所有 n 维复内积空间都同构，它们皆同构于酉空间 C^n。因此，n 维线性空间 R^n 及 C^n，当分别定义了实内积与复内积后，就成为 n 维欧氏空间与 n 维酉空间的典型代表，两者在应用中都很重要。今后在许多情况下提及到 R^n 与 C^n 时，都指它们已经是赋予了内积的线性空间，即欧氏空间与酉空间。

§7　正规矩阵

有一类矩阵 A，如对角形矩阵、实对称矩阵($A^T=A$)、实反对称矩阵($A^T=-A$)、厄米特矩阵* ($A^H=A$)、反厄米特矩阵*($A^H=-A$)、正交矩阵($A^TA=AA^T=E$)以及酉矩阵($A^HA=AA^H=E$)等，都有一个共同的性质：$A^HA=AA^H$。为了能够用统一的方法研究它们的相似标准形，我们引入正规矩阵的概念。

定义 1　设 $A\in C^{n\times n}$，且 $A^HA=AA^H$，则称 A 为**正规矩阵**。

容易验证，上面提到的几个特殊的矩阵都是正规矩阵。当然正规矩阵并非只是这些，例如

$$A=\begin{pmatrix}1 & -1\\ 1 & 1\end{pmatrix}$$

---

* 这里可算作是厄米特矩阵与反厄米特矩阵的正式定义，即是说，前者是满足条件 $A^H=A$ 的复数矩阵 A。同理理解后者。

这是一个正规矩阵，它不属于上述矩阵的任何一种。

在证明下面的重要定理1之前，先证明一个引理。

引理 若 $e_1=(a_1,a_2,\cdots,a_n)^T$ 是酉空间 C^n 的一个单位向量，则存在一个以 e_1 为第一个列向量的酉矩阵 Q。

证明 因 $e_1\neq\theta$，故 a_i 不全为零。设 $x=(x_1,x_2,\cdots,x_n)^T\in C^n$，且 $(x^T,e_1^T)=0$。由 C^n 内积定义，得

$$\bar{a}_1x_1+\bar{a}_2x_2+\cdots+\bar{a}_nx_n=0\ . \qquad (1)$$

这里 $\bar{a}_i$ 必不全为零。以 $x_1,x_2,\cdots,x_n$ 为未知数的线性方程(1)必有非零解，且解空间是 $n-1$ 维的。设 $\varepsilon_2,\varepsilon_3,\cdots,\varepsilon_n$ 是它的 $n-1$ 个线性无关的解向量(列向量)，设施行正交化、单位化后得到标准正交组 $e_2,e_3,\cdots,e_n$，显然它们仍然是线性方程(1)的解向量，从而都与 e_1 正交，所以

$$e_1,e_2,e_3,\cdots,e_n$$

是 C^n 的一组标准正交基，因此

$$Q=(e_1,e_2,e_3,\cdots,e_n)$$

是酉矩阵。证毕。

定理1 矩阵 $A\in C^{n\times n}$ 为正规矩阵的充要条件是：存在酉矩阵 Q，使得 A 酉相似于对角形矩阵，即

$$Q^HAQ=Q^{-1}AQ=\begin{bmatrix}\lambda_1&&&\\&\lambda_2&&\\&&\ddots&\\&&&\lambda_n\end{bmatrix} \qquad (1)$$

这里 $\lambda_1,\lambda_2,\cdots,\lambda_n$ 是 A 的特征值。

证明 充分性。设(1)式成立，则有

$$A=Q\begin{bmatrix}\lambda_1&&&\\&\lambda_2&&\\&&\ddots&\\&&&\lambda_n\end{bmatrix}Q^{-1},$$

$$A^H = Q\begin{bmatrix} \bar{\lambda}_1 & & & \\ & \bar{\lambda}_2 & & \\ & & \ddots & \\ & & & \bar{\lambda}_n \end{bmatrix}Q^{-1}.$$

于是有

$$AA^H = Q\begin{bmatrix} \lambda_1\bar{\lambda}_1 & & & \\ & \lambda_2\bar{\lambda}_2 & & \\ & & \ddots & \\ & & & \lambda_n\bar{\lambda}_n \end{bmatrix}Q^{-1}$$

$$= A^H A.$$

所以，A 是正规矩阵。

下面来证必要性。对于一阶矩阵，定理成立，设定理对 $n-1$ 阶矩阵已成立，现在证明对于 n 阶正规矩阵 A，定理的结论也能成立。

设 λ_1 是 A 的一个特征值，e_1 是 A 的属于 λ_1 的单位特征向量，$Ae_1=\lambda_1 e_1$。由上述引理，存在以 e_1 为第一个列向量的酉矩阵 Q_1。设

$$Q_1 = (e_1, e_2, \cdots, e_n),$$

由
$$E = Q_1^{-1}Q_1$$
$$= (Q_1^{-1}e_1, Q_1^{-1}e_2, \cdots, Q_1^{-1}e_n),$$

可得
$$Q_1^{-1}e_1 = (1, 0, \cdots, 0)^T.$$

又
$$Q_1^{-1}AQ_1 = Q_1^{-1}(Ae_1, Ae_2, \cdots, Ae_n)$$
$$= (\lambda_1 Q_1^{-1}e_1, Q_1^{-1}Ae_2, \cdots, Q_1^{-1}Ae_n)$$
$$= \begin{bmatrix} \lambda_1 & b_2 & \cdots & b_n \\ 0 & & & \\ \vdots & & B & \\ 0 & & & \end{bmatrix} = \begin{pmatrix} \lambda_1 & \beta \\ 0 & B \end{pmatrix},$$

其中 $\beta=(b_2, \cdots, b_n)$，又 B 是 $n-1$ 阶方阵。

设 $A_1=Q_1^{-1}AQ_1$，则有

$$\begin{aligned}A_1^H A_1 &=(Q_1^{-1}AQ_1)^H \cdot (Q_1^{-1}AQ_1)\\ &=Q_1^H A^H (Q_1^{-1})^H \cdot Q_1^{-1}AQ_1\\ &=Q_1^H A^H A Q_1 \qquad ((Q_1^{-1})^H = Q_1)\\ &=Q_1^{-1}AA^H Q_1; \quad (A^H A = AA^H)\end{aligned}$$

同理，有 $A_1A_1^H=Q_1^{-1}AA^HQ_1$。因此 $A_1^HA_1=A_1A_1^H$，亦即 A_1 是正规矩阵。又因

$$A_1^H A_1 = \begin{pmatrix}\lambda_1 & \beta\\ 0 & B\end{pmatrix}^H \begin{pmatrix}\lambda_1 & \beta\\ 0 & B\end{pmatrix} = \begin{pmatrix}\bar{\lambda}_1 & 0\\ \beta^H & B^H\end{pmatrix}\begin{pmatrix}\lambda_1 & \beta\\ 0 & B\end{pmatrix}$$

$$= \begin{pmatrix}\bar{\lambda}_1\lambda_1 & \bar{\lambda}_1\beta\\ \lambda_1\beta^H & \beta^H\beta + B^H B\end{pmatrix} \tag{1}$$

同理，有

$$A_1A_1^H = \begin{pmatrix}\lambda_1\bar{\lambda}_1 + \beta\beta^H & \beta B^H\\ B\beta^H & BB^H\end{pmatrix} \tag{2}$$

由于 $A_1^HA_1=A_1A_1^H$，故由(1)、(2)式右边比较左上角元素便有

$$\bar{\lambda}_1\lambda_1 = \lambda_1\bar{\lambda}_1 + \beta\beta^H,$$

因此 $\beta\beta^H=0$。亦即

$$(b_2,\cdots,b_n)\begin{bmatrix}\bar{b}_2\\ \vdots\\ \bar{b}_n\end{bmatrix} = b_2\bar{b}_2 + \cdots + b_n\bar{b}_n = 0 .$$

从而，$b_2=b_3=\cdots=b_n=0$，故 $\beta=(0,0,\cdots,0)$，即 β 是零向量。故由(1)、(2)两式又得

$$B^H B = BB^H,$$

即 B 是 $n-1$ 阶正规矩阵。由归纳法假设存在 $n-1$ 阶酉矩阵 C，使得

$$C^H BC = \begin{bmatrix} \lambda_2 & & & \\ & \lambda_3 & & \\ & & \ddots & \\ & & & \lambda_n \end{bmatrix}$$

令

$$Q_2 = \begin{pmatrix} 1 & 0 \\ 0 & C \end{pmatrix}$$

则 Q_2 是 n 阶酉矩阵。又 $Q=Q_1Q_2$ 也是酉矩阵，而且

$$\begin{aligned} Q^H AQ &= Q_2^H(Q_1^H AQ_1)Q_2 \\ &= \begin{pmatrix} 1 & 0 \\ 0 & C^H \end{pmatrix}\begin{pmatrix} \lambda_1 & 0 \\ 0 & B \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 0 & C \end{pmatrix} \\ &= \begin{pmatrix} \lambda_1 & 0 \\ 0 & C^H BC \end{pmatrix} \\ &= \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{bmatrix} \end{aligned}$$

又

$$\begin{aligned} |\lambda E - Q^H AQ| &= |Q^H(\lambda E - A)Q| \\ &= |\lambda E - A| \cdot |Q^H Q| \\ &= |\lambda E - A|, \end{aligned}$$

故 $Q^H AQ$ 的特征值 $\lambda_1, \lambda_2, \cdots, \lambda_n$ 也是 A 的特征值。故定理的必要性得证。证毕。

推论 1 设 A 是 n 阶正规矩阵，其特征值为 $\lambda_1, \lambda_2, \cdots, \lambda_n$，则

1) A 是厄米特(Hermite)矩阵的充要条件，是 A 的特征值全为实数；

2) A 是反厄米特矩阵的充要条件，是 A 的特征值为零或纯虚数；

3) A 是酉矩阵的充要条件，是 A 的每个特征值 λ_i 的模 $|\lambda_i|=1$。

证明

1) 因 A 是正规矩阵，由定理 1，存在酉矩阵 Q，使得

$$Q^H AQ = \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{bmatrix} \tag{1}$$

对这等式两边取共轭转置即得

$$Q^H A^H Q = \begin{bmatrix} \bar{\lambda}_1 & & & \\ & \bar{\lambda}_2 & & \\ & & \ddots & \\ & & & \bar{\lambda}_n \end{bmatrix} \tag{2}$$

若 A 为厄米特矩阵，即 $A^H=A$，则(2)式成为：

$$Q^H AQ = \begin{bmatrix} \bar{\lambda}_1 & & & \\ & \bar{\lambda}_2 & & \\ & & \ddots & \\ & & & \bar{\lambda}_n \end{bmatrix} \tag{3}$$

比较(1)式与(3)式，即得 $\lambda_i=\bar{\lambda}_i(i=1,2,\cdots,n)$。因此，$A$ 的特征值 $\lambda_1,\lambda_2,\cdots,\lambda_n$ 全为实数。

反之，若正规矩阵 A 的特征值全为实数，则有 $\lambda_i=\bar{\lambda}_i$，$i=1,2,\cdots,n$，这时，由于(1)式与(2)式的右边相等。所以左边也相等：

$$Q^H AQ = Q^H A^H Q\ 。$$

由于酉矩阵 Q 是可逆的，且 $Q^H=Q^{-1}$，故由上式易得 $A=A^H$，即 A 为厄米特矩阵。

2) 仿照 1)的证法便可证得。

3) 因 A 是正规矩阵，故上面的(1)式与(2)式成立。将(2)式右

乘到(1)式即得：

$$Q^H AQ \cdot Q^H A^H Q = \begin{bmatrix} \lambda_1\bar{\lambda}_1 & & & \\ & \lambda_2\bar{\lambda}_2 & & \\ & & \ddots & \\ & & & \lambda_n\bar{\lambda}_n \end{bmatrix}$$

因 $QQ^H = Q^H Q = E$，故有

$$Q^H AA^H Q = \begin{bmatrix} \lambda_1\bar{\lambda}_1 & & & \\ & \lambda_2\bar{\lambda}_2 & & \\ & & \ddots & \\ & & & \lambda_n\bar{\lambda}_n \end{bmatrix} \tag{4}$$

如果 A 为酉矩阵，则 $AA^H = E$，故(4)式变为

$$E = \begin{bmatrix} \lambda_1\bar{\lambda}_1 & & & \\ & \lambda_2\bar{\lambda}_2 & & \\ & & \ddots & \\ & & & \lambda_n\bar{\lambda}_n \end{bmatrix} \tag{5}$$

于是，有 $\lambda_i\bar{\lambda}_i = |\lambda_i|^2 = 1$，即 $|\lambda_i| = 1$，$(i = 1, 2, \cdots, n)$。

反过来，若 A 的每个特征值的模 $|\lambda_i| = 1$，则由(4)式便得

$$Q^H AA^H Q = E,$$

从而 $AA^H = E$。由此可得知 A 是酉矩阵。证毕。

注意，由于实对称矩阵是厄米特矩阵的特殊情形，故由推论 1 得知：实对称矩阵的特征值都是实数。同样地，由于正交矩阵是酉矩阵的特例，故正交矩阵的特征值的模均等于 1。

推论 2 厄米特矩阵 $A \in C^{n\times n}$ 的任二不同特征值 λ, μ，所对应的特征向量 $x, y \in C^n$ 是正交的。

证明 因 $A^H = A$，故由

$$Ax = \lambda x, \quad Ay = \mu y \quad (x, y \neq \theta)$$

的后一式取共轭转置便得：

$$y^H A = \mu y^H。$$

两边右乘以 x，即得

$$y^H A x = \mu y^H x 。$$

但 $y^H A x = y^H(\lambda x) = \lambda y^H x$，故由上式便有

$$(\lambda - \mu) y^H x = 0 。$$

但 $\lambda \neq \mu$，所以 $y^H x = 0$，即 $(x, y) = 0$，因此 x, y 正交。证毕。

□注： 对于实矩阵虽然亦可以定义正规矩阵，其定义为：若 $A \in R^{n \times n}$，且 $A^T A = A A^T$，则称 A 为实正规矩阵。但此时一般不能断言 A 正交相似于对角形矩阵，即不要以为把定理 1 的"酉"字改为"正交"两字就可把结论搬到实正规矩阵上来。这情况相当复杂，这里不作讨论。但是实正规矩阵的一些特殊情形（如反对称矩阵，正定矩阵）的某些结论，在研究陀螺系统的运动情况、弹性系统的振动（经有限元处理后）及一般保守力系下的小振动等问题上却是很有用的。□

§8 厄米特二次型

我们在上面已经知道，若 $A \in C^{n \times n}$，且 $A^H = A$（即 $\bar{a}_{ji} = a_{ij}$ 对所有 $i, j = 1, 2, \cdots, n$ 成立），则 A 叫做厄米特矩阵。它是实对称矩阵的一种推广。由 $\bar{a}_{ii} = a_{ii}(i = 1, 2, \cdots, n)$ 得知厄米特矩阵 A 的主对角线上的元素全是实数。

定义 1 若 $x = (x_1, x_2, \cdots, x_n)^T \in C^n$，又 $A = (a_{ij}) \in C^{n \times n}$ 为厄米特矩阵，则二次型

$$f(x) = \sum_{i,j=1}^{n} a_{ij} \bar{x}_i x_j,$$

即

$$f(x) = x^H A x,$$

称为厄米特二次型。A 的秩称为这二次型的秩。

定理 1 厄米特二次型 $f=x^H Ax$ 经满秩线性变换 $x=Cy$ ($C\in C^{n\times n}$, $|C|\neq 0$),仍为厄米特二次型,且秩不变。

证明 因

$$
\begin{aligned}
f(x) &= x^H Ax \\
&= (Cy)^H A(Cy) \\
&= y^H(C^H AC)y = y^H By \ .
\end{aligned}
$$

这里 $B=C^H AC$ 也是厄米特矩阵,因为

$$
\begin{aligned}
B^H &= (C^H AC)^H = C^H A^H (C^H)^H \\
&= C^H AC = B \ .
\end{aligned}
$$

又显然 B 的秩等于 A 的秩,故 $f_1(y)=y^H By$ 也是厄米特二次型,且秩与 $f(x)$ 相同。证毕。

因为每个厄米特二次型完全被它的系数矩阵——厄米特矩阵所确定,所以研究厄米特二次型同研究厄米特矩阵是相当的。厄米特二次型经过变量代换,系数矩阵的变化如定理 1 证明中所给出的,我们称矩阵 A 与矩阵 $B=C^H AC$ 是厄米特相合的。所以,一个厄米特二次型经满秩线性变换 $x=Cy$ 化为标准形

$$
f_1(y) = \sum_{i=1}^{n} b_i \bar{y}_i y_i
$$

的问题,就相当于矩阵 A 与矩阵

$$
B = \begin{bmatrix} b_1 & & & \\ & b_2 & & \\ & & \ddots & \\ & & & b_n \end{bmatrix}
$$

厄米特相合的问题。

由上节定理 1 及推论 1,即知:

每个厄米特矩阵 $A\in C^{n\times n}$,都存在酉矩阵 $Q\in C^{n\times n}$,使得 A 酉相似于对角形矩阵:

$$Q^H AQ = Q^{-1}AQ = \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{bmatrix}$$

也就是 A 厄米特相合于对角形矩阵，这里 $\lambda_i(i=1,2,\cdots,n)$ 都是实数，它们是 A 的全部特征值。

对上述的酉矩阵 Q，作空间 C^n 的线性变换

$$x = Qy,\quad (y \text{ 为 } C^n \text{ 中列向量})$$

现证明这是个酉变换．设 y,z 为 C^n 中任二列向量，则由酉空间 C^n 内积定义及 $Q^HQ=E$ 可得：

$$\begin{aligned}(Qy,Qz) &= (Qz)^H Qy \\ &= z^H Q^H Qy \\ &= z^H y = (y,z)\ .\end{aligned}$$

这就是所要证明的。

转换成二次型的语言，我们便有：

定理 2 每个厄米特二次型 $f=x^HAx$，都可用某个酉变换 $x=Qy$（Q 是酉矩阵），使其化为标准形：

$$f = \lambda_1\bar{y}_1y_1 + \lambda_2\bar{y}_2y_2 + \cdots + \lambda_n\bar{y}_ny_n,$$

其中 $\lambda_1,\lambda_2,\cdots,\lambda_n$ 是 A 的特征值。

例 1 化厄米特二次型

$$f = \bar{x}\,x_1 - i\bar{x}_2x_1 + \bar{x}_3x_1 + i\bar{x}_1x_2 + 2i\bar{x}_3x_2 + \bar{x}_1x_3 - 2i\bar{x}_2x_3$$

为标准形。

解 这二次型的矩阵为

$$A = \begin{bmatrix} 1 & i & 1 \\ -i & 0 & -2i \\ 1 & 2i & 0 \end{bmatrix}$$

又 $|\lambda E-A|=\lambda^3-\lambda^2-6\lambda$，故 A 有相异特征值：

$$\lambda_1 = 0,\lambda_2 = 3,\lambda_3 = -2.$$

相应的特征向量为：

$$\varepsilon_1=(2,i,-1)^T,\quad \varepsilon_2=(i,1,i)^T,\quad \varepsilon_3=(0,1,-i)^T.$$

因 A 是厄米特矩阵，由上节的推论 2，这些特征向量是两两正交的。再将它们化为单位向量：

$$e_1=\frac{1}{\sqrt{6}}\varepsilon_1,\quad e_2=\frac{1}{\sqrt{3}}\varepsilon_2,\quad e_3=\frac{1}{\sqrt{2}}\varepsilon_3.$$

注意，这里计算向量的长度时，是把这向量视为酉空间 C^3 的向量并按 C^3 的内积来计算的，如

$$|\varepsilon_1|=\sqrt{(\varepsilon_1,\varepsilon_1)}=\sqrt{2^2+i\cdot(-i)+(-1)^2}=\sqrt{6}.$$

以 e_1，e_2，e_3 作为列向量构成矩阵

$$Q=\begin{bmatrix}\frac{2}{\sqrt{6}} & \frac{i}{\sqrt{3}} & 0\\ \frac{i}{\sqrt{6}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{2}}\\ \frac{-1}{\sqrt{6}} & \frac{i}{\sqrt{3}} & \frac{-i}{\sqrt{2}}\end{bmatrix}$$

为酉矩阵。作酉变换 $x=Qy$ $(y=(y_1,y_2,y_3)^T)$，则有标准形：

$$f=3\bar{y}_2y_2-2\bar{y}_3y_3.$$

注意，由定理 2 可知，对每个厄米特二次型 $f=x^HAx$，都存在满秩线性变换 $x=Qy$，使其化为标准形 $f=\sum_{i=1}^{n}\lambda_i|y_i|^2$，且系数 λ_i 为 A 的特征值，$i=1,2,\cdots,n$。

定义 2 若对任一 $x\neq\theta$，厄米特二次型 $f=x^HAx$ 恒为正（负）数，则称它是正（负）定的，这时厄米特矩阵 A 也称为正（负）定的；若 f 恒不为负（正）数，则 f 叫做半正（负）定的，相应地 A 也叫做半正（负）定的。

定理 3 厄米特二次型 $f=x^HAx$ 为正定的充要条件，是 A 的特征值全为正数。

证明 由定理 2，经过满秩线性变换 $x=Qy$（Q 为酉矩阵），f 化为标准形

$$f = \sum_{i=1}^{n} \lambda_i |y_i|^2,$$

$\lambda_1,\lambda_2,\cdots,\lambda_n$ 是 A 的特征值，当 $x\neq\theta$，则 $y\neq\theta$，故当 A 的特征值全为正数时，便有 $f>0$。

必要性。设 f 为正定，假如在上面的标准形中，有一个 $\lambda_i\leqslant 0$，如果取非零向量

$$x = Q(0,\cdots,0,1,0,\cdots,0)^T \quad (\text{仅第 } i \text{ 个分量为 } 1)$$

时，则有 $f=\lambda_i\leqslant 0$，这与 f 为正定相矛盾．证毕．

类似地可以证明：

厄米特二次型 $f=x^H Ax$ 为半正定的充要条件，是 A 的特征值 $\lambda_i\geqslant 0 \quad (i=1,2,\cdots,n)$。

不难证明：

定理 4 若 A 为厄米特矩阵，则下列两个条件中的任何一个，都是 A 为正定矩阵的充要条件：

（1）存在满秩矩阵 C，使得 $C^H AC=E$；

（2）存在满秩矩阵 B，使得 $A=B^H B$。

此外，与实二次型的情形一样，用归纳法可以证明（证略）下述定理：

定理 5 厄米特二次型 $f=x^H Ax$ 为正定的充要条件，是 A 的各阶顺序主子式大于零．

最后我们来证明一个关于把两个二次型同时化为标准形的定理，它在微振动理论中是有用的．

定理 6 设

$$f(x_1,x_2,\cdots,x_n) = x^H Ax,$$

$$g(x_1,x_2,\cdots,x_n) = x^H Bx$$

是两个厄米特二次型，且 $g=x^H Bx$ 正定．则存在满秩线性变换

$x=Cy$,使这两个二次型同时化为标准形：

$$f=\sum_{k=1}^{n}\lambda_k|y_k|^2,\quad g=\sum_{k=1}^{n}|y_k|^2.$$

这里 $\lambda_1,\lambda_2,\cdots,\lambda_n$ 是方程 $|\lambda B-A|=0$ 的根。

证明 因 B 为正定厄米特矩阵，故存在满秩矩阵 C_1，使得 $C_1^H BC_1=E$。

对于 C_1，我们考察矩阵 $D=C_1^H AC_1$。由于 $A^H=A$，所以

$$D^H=(C_1^H AC_1)^H=C_1^H A^H(C_1^H)^H=C_1^H AC_1,$$

即 D 也是厄米特矩阵。故又有酉矩阵 C_2，使得

$$C_2^H DC_2=C_2^H(C_1^H AC_1)C_2=\begin{bmatrix}\lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n\end{bmatrix}$$

令 $C=C_1C_2$，则

$$\begin{aligned}C^H BC&=(C_1C_2)^H B(C_1C_2)=C_2^H(C_1^H BC_1)C_2\\&=C_2^H EC_2=C_2^H C_2=E\quad(C_2\text{ 是酉矩阵}),\\C^H AC&=C_2^H(C_1^H AC_1)C_2\\&=\begin{bmatrix}\lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n\end{bmatrix}=\mathrm{diag}(\lambda_1,\lambda_2,\cdots,\lambda_n).\end{aligned}$$

故经满秩线性变换 $x=Cy$，即有：

$$f=\sum_{k=1}^{n}\lambda_k|y_k|^2,\quad g=\sum_{k=1}^{n}|y_k|^2.$$

又

$$\begin{aligned}|C^H(\lambda B-A)C|&=|\lambda(C^H BC)-C^H AC|\\&=|\lambda E-\mathrm{diag}(\lambda_1,\lambda_2,\cdots,\lambda_n)|,\end{aligned}$$

即：

$$\overline{|C|}\cdot|C|\cdot|\lambda B-A|=\begin{vmatrix}\lambda-\lambda_1 & & & \\ & \lambda-\lambda_2 & & \\ & & \ddots & \\ & & & \lambda-\lambda_n\end{vmatrix}$$

由于$|C|\neq0$，故常数

$$a=\overline{|C|}\cdot|C|\neq0。$$

因此

$$|\lambda B-A|=\frac{1}{a}(\lambda-\lambda_1)(\lambda-\lambda_2)\cdots(\lambda-\lambda_n)。$$

即是说，前面出现的数$\lambda_1,\lambda_2,\cdots,\lambda_n$是方程

$$|\lambda B-A|=0$$

的根。证毕。

定理 6 中的$\lambda_1,\lambda_2,\cdots,\lambda_n$称为矩阵$A$相对于矩阵$B$的广义特征值，当$B=E$时，也就是$A$的特征值。

作为特例，当定理 6 中的两个二次型都为实二次型，且至少有一个为正定，比如g为正定，则定理当然也是成立的。

§9 力学系统的小振动

现考察接近于它的平衡位置的，有n个自由度的不变力学系统的自由振动。这系统对平衡位置的偏差是用独立的广义坐标q_1，$q_2,\cdots,q_n$来给出的，而且它的平衡位置对应于$q_i=0$ $(i=1,2,\cdots,n)$。这时，系统的动能可表示为广义速度$\dot{q}_i(i=1,2,\cdots,n)$的二次型

$$T=\sum_{i,j=1}^{n}b_{ij}(q_1,q_2,\cdots,q_n)\dot{q}_i\dot{q}_j。$$

按照$q_1,q_2,\cdots,q_n$的次数的次序来排列$b_{ij}(q_1,q_2,\cdots,q_n)$的系

数，

$$b_{ij}(q_1,q_2,\cdots,q_n)=b_{ij}+\cdots,$$

而且（由于 $q_1,q_2,\cdots,q_n$ 为很微小的偏差）只保留常数项 b_{ij}，则有：

$$T=\sum_{i,j=1}^{n}b_{ij}\dot{q}_i\dot{q}_j \quad （每个\ b_{ij}=b_{ji}）. \tag{1}$$

由于动能常是正的，且只对于零速度 $\dot{q}_1=\dot{q}_2=\cdots=\dot{q}_n=0$ 时始变为零，故二次型(1)是正定的。

又位能是坐标的函数 $U=U(q_1,q_2,\cdots,q_n)$。不失一般性，取 $U_0=U(0,0,\cdots,0)=0$。同样地按 $q_1,q_2,\cdots,q_n$ 的次数的次序来裂分位能，得到：

$$U=\sum_{i=1}^{n}a_iq_i+\sum_{i,j=1}^{n}a_{ij}q_iq_j+\cdots.$$

因为在平衡位置位能常有固定的值，所以

$$a_i=\left(\frac{\partial U}{\partial q_i}\right)_0=0 \quad (i=1,2,\cdots,n)$$

而只保留 $q_i(i=1,2,\cdots,n)$ 的二次项，于是有

$$U=\sum_{i,j=1}^{n}a_{ij}q_iq_j \quad (a_{ij}=a_{ji}) \tag{2}$$

因而，动能与位能给出了两个二次型(1)与(2)，且第一个是正定的，故由上节定理 6，存在满秩线性变换

$$q=C\eta, \quad （这时\ \dot{q}=C\dot{\eta}，见第四章\ §6）$$

同时化这两个二次型为标准形：

$$T=\sum_{k=1}^{n}\dot{\eta}_k^2, \quad U=\sum_{k=1}^{n}\lambda_k\eta_k^2 \quad .$$

这里，

$$q=(q_1,q_2,\cdots,q_n)^T, \quad \eta=(\eta_1,\eta_2,\cdots,\eta_n)^T,$$

$$C=(c_{ij})_{n\times n}.$$

习 题 二

1. 设 $x=(\xi_1,\xi_2)$，$y=(\eta_1,\eta_2)$ 是二维实线性空间 R^2 的任二向量，问 R^2 对以下定义的内积是否构成欧氏空间：

(1) $(x,y)=\xi_1\eta_1+\xi_2\eta_2+1$；

(2) $(x,y)=\xi_1\eta_1-\xi_2\eta_2$；

(3) $(x,y)=3\xi_1\eta_1+5\xi_2\eta_2$。

2. 设 V 是实数域 R 上的 n 维线性空间，$e_1,e_2,\cdots,e_n$ 是 V 的一组基，对 V 中任二向量

$$x=\sum_{i=1}^{n}\xi_i e_i,\quad y=\sum_{i=1}^{n}\eta_i e_i,$$

规定

$$(x,y)=\sum_{i=1}^{n}i\xi_i\eta_i \quad .$$

证明 (x,y) 是 V 中一种内积，从而 V 对此内积作成一欧氏空间。

3. 在欧氏空间 R^4 中，求一单位向量与下列三个向量正交：

$$(1,1,-1,1),(1,-1,-1,1),(2,1,1,3)\ .$$

4. 设 $\alpha_1,\alpha_2,\cdots,\alpha_m$ 是欧氏空间 V 的一组向量，证明这组向量线性无关的充要条件是行列式

$$\begin{vmatrix} (\alpha_1,\alpha_1) & (\alpha_1,\alpha_2) & \cdots & (\alpha_1,\alpha_m) \\ (\alpha_2,\alpha_1) & (\alpha_2,\alpha_2) & \cdots & (\alpha_2,\alpha_m) \\ \cdots & \cdots & \cdots & \cdots \\ (\alpha_m,\alpha_1) & (\alpha_m,\alpha_2) & \cdots & (\alpha_m,\alpha_m) \end{vmatrix}\neq 0\ .$$

5. 证明：对任意实数 $a_1,a_2,\cdots,a_n$，下列不等式成立：

$$\sum_{i=1}^{n}|a_i|\leqslant\sqrt{n\sum_{i=1}^{n}a_i^2}\ .$$

6. 设 e_1,e_2,e_3 是三维欧氏空间的一组标准正交基，证明：

$$\alpha_1=\frac{1}{3}(2e_1+2e_2-e_3),$$

$$\alpha_2=\frac{1}{3}(2e_1-e_2+2e_3),$$

$$a_3 = \frac{1}{3}(e_1 - 2e_2 - 2e_3)$$

也是一组标准正交基。

7. 求齐次线性方程组

$$\begin{cases} 2x_1 + x_2 - x_3 + x_4 - 3x_5 = 0 \\ x_1 + x_2 - x_3 + x_5 = 0 \end{cases}$$

的解空间(作为 R^5 的子空间)的一组标准正交基。

8. 设 V 是 n 维欧氏空间，α 为 V 中一个取定的非零向量，证明：

(1) $V_1 = \{x \mid (x, \alpha) = 0, x \in V\}$ 是 V 的子空间；

(2) $\dim V_1 = n - 1$。

9. 设 $\alpha_1, \alpha_2, \alpha_3$ 是三维欧氏空间 V 的一组标准正交基，求 V 的一个正交变换 T，使得

$$\begin{cases} T\alpha_1 = \frac{2}{3}\alpha_1 + \frac{2}{3}\alpha_2 - \frac{1}{3}\alpha_3 \\ T\alpha_2 = \frac{2}{3}\alpha_1 - \frac{1}{3}\alpha_2 + \frac{2}{3}\alpha_3. \end{cases}$$

10. 证明：欧氏空间中两个正交变换的乘积也是正交变换；正交变换的逆变换也是正交变换。

11. 证明：如果一个上三角矩阵

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ 0 & a_{22} & \cdots & a_{2n} \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & a_{nn} \end{bmatrix}$$

是正交矩阵，则 A 必为对角形矩阵，且主对角线上的元素 $a_{ii} = \pm 1 \quad (i = 1, 2, \cdots, n)$。

12. 证明：n 阶方阵 A 为酉矩阵的充要条件，是对任何行向量 $x \in C^n$，都有 $|xA| = |x|$。

13. 设 P, Q 各为 m 阶及 n 阶方阵，证明：若 $m+n$ 阶方阵

$$A = \begin{pmatrix} P & B \\ 0 & Q \end{pmatrix}$$

是酉矩阵，则 P, Q 也是酉矩阵，且 B 是零矩阵。

14. 设 A, B 均为厄米特矩阵，证明：AB 为厄米特矩阵的充要条件是 AB

$=BA$。

15. 证明：任一复数方阵都可以表示成厄米特矩阵与反厄米特矩阵之和。

16. 试求一酉矩阵 P，使 $P^{-1}AP$ 为对角形：

(1) $A=\begin{bmatrix}-1 & i & 0\\ -i & 0 & -i\\ 0 & i & -1\end{bmatrix}$　　(2) $A=\begin{bmatrix}0 & i & 1\\ -i & 0 & 0\\ 1 & 0 & 0\end{bmatrix}$

这两个矩阵是正规矩阵吗？

17. 证明：两个正规矩阵相似（酉等价）的充要条件，是特征多项式相同。

18. 若两个正规矩阵可交换，证明它们的乘积也是正规矩阵。

第三章　矩阵的标准形与若干分解形式

本章集中讨论数字矩阵、多项式矩阵、有理分式矩阵的标准形及矩阵的若干分解形式．这是矩阵理论中一个内容广泛而又十分重要的部分，它在许多领域中都有重要的应用．在这里主要是结合控制系统方面的应用来介绍多项式矩阵及有理分式矩阵的某些内容的，而矩阵的 QR 分解、奇异值分解等在数值分析及其他领域都有重要应用．对这些应用性较强的材料，我们只能给出一个导引性的讨论，而不能作比这更深入、更详尽的阐述了．

§1　矩阵的相似对角形

本节主要研究：一个 n 阶矩阵能够相似于对角形矩阵的充要条件是什么？解决这个问题而得到的结果同后面§3的主要结果一起，构成矩阵理论的主要工具之一．

设 V 是复数域 C 上的 n 维线性空间，T 是 V 的一个线性变换．又 $e_1,e_2,\cdots,e_n$ 与 $\varepsilon_1,\varepsilon_2,\cdots,\varepsilon_n$ 是 V 的两组基，从第一组基到第二组基的过渡矩阵是 P．则线性变换 T 在这两组基下的矩阵 A 与 B 相似，即

$$B = P^{-1}AP .$$

自然会问：矩阵 A 可否相似于一个对角形矩阵？换言之，是否可以适当地选取第二组基 $\varepsilon_1,\varepsilon_2,\cdots,\varepsilon_n$，使得线性变换 T 在这组基下的矩阵 B 是个对角形矩阵呢？下面，我们将逐步弄清这个问题．

首先设想矩阵 A 能相似于一个对角形矩阵，即设

$$P^{-1}AP = \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{bmatrix} \quad (1)$$

因而有

$$AP = P\begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{bmatrix} \quad (2)$$

若把 P 写成分块矩阵：

$$P = (X_1, X_2, \cdots, X_n),$$

这里 $X_1, X_2, \cdots, X_n$ 代表 P 的 n 个列向量。应用矩阵乘法规则，容易验证

$$AP = (AX_1, AX_2, \cdots, AX_n),$$

$$P\begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{bmatrix} = (\lambda_1 X_1, \lambda_2 X_2, \cdots, \lambda_n X_n).$$

故由(2)式可得

$$AX_i = \lambda_i X_i \quad (i = 1, 2, \cdots, n), \quad (3)$$

或

$$(\lambda_i E - A)X_i = \theta \quad (i = 1, 2, \cdots, n).$$

这说明，若 A 能与对角形矩阵相似（见(1)式），则可逆矩阵 $P = (X_1, X_2, \cdots, X_n)$ 的每个列向量（非零向量）X_i 都满足(3)式。

读者已经知道，若 A 为 n 阶矩阵，X 是 n 维列向量，则满足方程 $AX = \lambda X$ 的非零向量 X 称为矩阵 A 的对应于（或属于）特征值 λ 的特征向量。因此，上述讨论说明：当 A 与对角形矩阵相似时，这对角形矩阵主对角线上的元素 $\lambda_i (i = 1, 2, \cdots, n)$ 都是 A 的特征值，

而矩阵 P 的 n 个列向量 $X_1,X_2,\cdots,X_n$ 是 A 的 n 个线性无关的特征向量(因 $|P|\neq 0$,故 P 的 n 个列向量线性无关)。反过来,若 A 有 n 个线性无关的特征向量 $X_1,X_2,\cdots,X_n$,相应的特征值分别为 λ_1, $\lambda_2,\cdots,\lambda_n$,即

$$AX_1=\lambda_1X_1, AX_2=\lambda_2X_2,\cdots,AX_n=\lambda_nX_n,$$

则取 $P=(X_1,X_2,\cdots,X_n)$。显然,P 为满秩矩阵。再由

$$AP=(AX_1,\cdots,AX_n)=(\lambda_1X_1,\cdots,\lambda_nX_n)$$

$$=(X_1,X_2,\cdots,X_n)\begin{bmatrix}\lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n\end{bmatrix}$$

$$=P\begin{bmatrix}\lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n\end{bmatrix},$$

可得

$$P^{-1}AP=\begin{bmatrix}\lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n\end{bmatrix}$$

即 A 与对角形矩阵相似。

这样,我们得到了下述定理。

定理 1 设 $A\in C^{n\times n}$,则 A 能与对角形矩阵相似的充要条件,是 A 有 n 个线性无关的特征向量。

因为属于 A 的不同特征值的特征向量线性无关,所以我们有:

定理 2(充分条件) 若 n 阶矩阵 A 有 n 个不同的特征值,则 A 可与对角形矩阵相似。

这个充分条件，在许多情况下使用起来是很方便的。

如何求矩阵的特征值与特征向量，读者是熟悉的，我们只作如下的简短说明。

上面已说过，若 A 为 n 阶矩阵，X 是 n 维列向量，则满足方程 $AX=\lambda X$ 的非零向量 X 称为 A 的属于特征值 λ 的特征向量。我们知道，$AX=\lambda X$ 亦即

$$(\lambda E - A)X = \theta$$

有非零解 $X\neq\theta$ 的充要条件，是行列式

$$|\lambda E - A| = 0\ . \qquad (4)$$

所以，A 的特征值 λ 应从方程(4)求出，它称为矩阵 A 的特征方程。而矩阵

$$\lambda E - A = \begin{bmatrix} \lambda - a_{11} & -a_{12} & \cdots & -a_{1n} \\ -a_{21} & \lambda - a_{22} & \cdots & -a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ -a_{n1} & -a_{n2} & \cdots & \lambda - a_{nn} \end{bmatrix}$$

称为 A 的特征矩阵，又多项式

$$f(\lambda) = |\lambda E - A| = \lambda^n + \alpha_1\lambda^{n-1} + \cdots + \alpha_i\lambda^{n-i} + \cdots + \alpha_n$$

称为 A 的特征多项式，这里 $\alpha_1=-\sum_{i=1}^{n}a_{ii}=-\mathrm{tr}(A)$，$\alpha_n=(-1)^n\cdot|A|$，$\alpha_i$ 是 A 的所有 i 阶主子式的和与 $(-1)^i$ 的乘积。在复数域中，n 阶矩阵 A 的特征多项式必有 n 个根(特征根)。又 $\mathrm{tr}(A)$ 叫 A 的迹。

易知相似矩阵有相同的特征多项式，从而有相同的特征值。

属于矩阵 A 的同一个特征值 λ_0 的所有特征向量连同零向量一起，构成一个线性空间 V_{λ_0}，称为 A 的特征子空间。特征子空间 V_{λ_0} 的维数不超过特征根 λ_0 的重数(证略)。

若 T 是数域 P 上 n 维线性空间 V 的线性变换，则满足 $Tx=\lambda x$ 的非零向量 x 称为线性变换 T 的属于特征值 λ 的特征向量，由线性变换与矩阵的对应关系，便知这与 $Ax=\lambda x$ 完全相当(设 T 与 A

对应)。因此,当线性变换 T 有 n 个线性无关的特征向量时,只要选取这组向量为一组基,则显然 T 在这组基下的矩阵是对角形矩阵,反过来,若 T 在某组基 $e_1,e_2,\cdots,e_n$ 下的矩阵是对角形矩阵 diag $(\lambda_1,\lambda_2,\cdots,\lambda_n)$,从而 $Te_i=\lambda_i e_i(i=1,2,\cdots,n)$,因此 $e_1,e_2,\cdots,e_n$ 是 T 的 n 个线性无关的特征向量. 其实,这就是定理 1 的另一种说法及证法。这证法简便得多了。

例 1 研究下列矩阵能否与对角形矩阵相似:

1) $A=\begin{bmatrix}5&6&-3\\-1&0&1\\1&2&-1\end{bmatrix}$ 2) $A=\begin{bmatrix}1&2&2\\2&1&2\\2&2&1\end{bmatrix}$

3) $A=\begin{bmatrix}3&1&0\\-4&-1&0\\4&-8&-2\end{bmatrix}$

解 1) 因 A 的特征多项式为

$$|\lambda E-A|=(\lambda-2)(\lambda^2-2\lambda-2),$$

因而 A 有三个不同的特征值:

$$\lambda_1=2,\lambda_2=1+\sqrt{3},\lambda_3=1-\sqrt{3}。$$

由定理 2 可知 A 能与对角形矩阵相似。又求得相应的三个特征向量为:

$$X_1=\begin{bmatrix}-2\\1\\0\end{bmatrix}\quad X_2=\begin{bmatrix}3\\-1\\2-\sqrt{3}\end{bmatrix}\quad X_3=\begin{bmatrix}3\\-1\\2+\sqrt{3}\end{bmatrix}$$

它们是属于三个不同特征值的,所以是线性无关的。以它们为列向量作矩阵

$$P=\begin{bmatrix}-2&3&3\\1&-1&-1\\0&2-\sqrt{3}&2+\sqrt{3}\end{bmatrix}$$

并求得

$$P^{-1}=\begin{bmatrix} -1 & -3 & 0 \\ \frac{1}{2}+\frac{\sqrt{3}}{3} & 1+\frac{2\sqrt{3}}{3} & -\frac{\sqrt{3}}{6} \\ \frac{1}{2}-\frac{\sqrt{3}}{3} & 1-\frac{2\sqrt{3}}{3} & \frac{\sqrt{3}}{6} \end{bmatrix}$$

且有

$$P^{-1}AP=\begin{bmatrix} 2 & & \\ & 1+\sqrt{3} & \\ & & 1-\sqrt{3} \end{bmatrix}$$

2）特征多项式为

$$|\lambda E-A|=(\lambda+1)^2(\lambda-5).$$

故特征值为

$$\lambda_1=\lambda_2=-1(\text{二重根}),\lambda_3=5.$$

又求得属于特征值-1的两个线性无关的特征向量为

$$X_1=(1,0,-1)^T, X_2=(0,1,-1)^T.$$

而特征值$\lambda_3=5$对应的一个特征向量为：

$$X_3=(1,1,1)^T.$$

因而有

$$P=\begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ -1 & -1 & 1 \end{bmatrix} \qquad P^{-1}=\frac{1}{3}\begin{bmatrix} 2 & -1 & -1 \\ -1 & 2 & -1 \\ 1 & 1 & 1 \end{bmatrix}$$

从而

$$P^{-1}AP=\begin{bmatrix} -1 & & \\ & -1 & \\ & & 5 \end{bmatrix}$$

3）A的特征多项式为

$$|\lambda E-A|=(\lambda-1)^2(\lambda+2),$$

特征值为$\lambda_1=\lambda_2=1,\lambda_3=-2$。而对应于特征值1的一切特征向量

为

$$X = K\begin{bmatrix}3\\-6\\20\end{bmatrix} \qquad K \neq 0.$$

又对应于特征值-2的一切特征向量为

$$Y = K_1\begin{bmatrix}0\\0\\1\end{bmatrix} \qquad K_1 \neq 0.$$

所以不存在三个线性无关的特征向量，从而 A 不能与对角形矩阵相似。

例 2 设

$$A = \begin{bmatrix}4 & 6 & 0\\-3 & -5 & 0\\-3 & -6 & 1\end{bmatrix}$$

求 A 的相似对角形及 A^{100}。

解 由 $|\lambda E - A| = (\lambda-1)^2(\lambda+2)$，得

$$\lambda_1 = -2, \quad \lambda_2 = 1 \quad （二重根）.$$

求得对应于 λ_1 的一个特征向量 $X_1 = (-1,1,1)^T$，及对应于二重根 $\lambda_2 = 1$ 的两个线性无关特征向量为

$$X_2 = (-2,1,0)^T, X_3 = (0,0,1)^T.$$

因此可得

$$P = \begin{bmatrix}-1 & -2 & 0\\1 & 1 & 0\\1 & 0 & 1\end{bmatrix}, \qquad P^{-1} = \begin{bmatrix}1 & 2 & 0\\-1 & -1 & 0\\-1 & -2 & 1\end{bmatrix}$$

因而有

$$P^{-1}AP = \begin{bmatrix}-2 & & \\ & 1 & \\ & & 1\end{bmatrix} \tag{5}$$

注意，若取

$$P_1 = \begin{bmatrix} -1 & 0 & -2 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix}$$

则亦有

$$P_1^{-1}AP_1 = \begin{bmatrix} -2 & & \\ & 1 & \\ & & 1 \end{bmatrix}$$

可见 P 不是唯一的。

现来计算 A^{100}。由(5)式得

$$A = P\begin{bmatrix} -2 & & \\ & 1 & \\ & & 1 \end{bmatrix}P^{-1}$$

因此易知

$$\begin{aligned} A^{100} &= P\begin{bmatrix} -2 & & \\ & 1 & \\ & & 1 \end{bmatrix}^{100} P^{-1} \\ &= \begin{bmatrix} -1 & -2 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}\begin{bmatrix} 2^{100} & & \\ & 1 & \\ & & 1 \end{bmatrix}\begin{bmatrix} 1 & 2 & 0 \\ -1 & -1 & 0 \\ -1 & -2 & 1 \end{bmatrix} \\ &= \begin{bmatrix} -2^{100}+2 & -2^{101}+2 & 0 \\ 2^{100}-1 & 2^{101}- & 0 \\ 2^{100}-1 & 2^{101}-2 & 1 \end{bmatrix}. \end{aligned}$$

§2 矩阵的约当标准形

由上节看到，并非每个矩阵都可以相似于对角形矩阵的。当矩阵 $A=(a_{ij})\in C^{n\times n}$不能和对角形矩阵相似时，能否找到一个构造比较简单的分块对角矩阵和它相似呢？当我们在复数域 C 内考虑这个问题时，这样的矩阵确实是存在的，这就是约当(Jordan)形矩阵，它称为矩阵 A 的约当标准形。在矩阵分析及其应用中，矩阵的约当标准形都是重要工具，但其理论推导，往往都十分繁复，所以我们在这里只作扼要介绍。

设 $A=(a_{ij})\in C^{n\times n}$，$\lambda E-A$ 是 A 的特征矩阵(见上节)，暂用 $A(\lambda)$来记它。

定义 1 $A(\lambda)$中所有非零的 k 级子式的首项(最高次项)系数为 1 的最大公因式 $D_k(\lambda)$称为 $A(\lambda)$的一个 k 级行列式因子，$k=1,2,\cdots,n$

由定义可知：$D_n(\lambda)=|\lambda E-A|$。又因为 $D_{k-1}(\lambda)$能整除每个 $k-1$ 级子式，从而可整除每个 k 级子式(将 k 级子式按一行或一列展开即知)，因此 $D_{k-1}(\lambda)$能整除 $D_k(\lambda)$，并记为 $D_{k-1}(\lambda)|D_k(\lambda)$ $(k=2,3,\cdots,n)$。

定义 2 下列的 n 个多项式

$$d_1(\lambda)=D_1(\lambda),d_2(\lambda)=\frac{D_2(\lambda)}{D_1(\lambda)},d_3(\lambda)=\frac{D_3(\lambda)}{D_2(\lambda)},$$

$$\cdots,d_k(\lambda)=\frac{D_k(\lambda)}{D_{k-1}(\lambda)},\cdots,d_n(\lambda)=\frac{D_n(\lambda)}{D_{n-1}(\lambda)},$$

称为 $A(\lambda)$的不变因式。把每个次数大于零的不变因式分解为互不相同的一次因式的方幂的乘积*，所有这些一次因式的方幂(相同

---

* 我们是在复数域内讨论的，这样的分解是可能的。

的必须按出现次数计算)，称为 $A(\lambda)$的初级因子。

由于这里的 $A(\lambda)=\lambda E-A$ 完全由 A 决定，所以这里 $A(\lambda)$的不变因式及初级因子，也常称为矩阵 A 的不变因式及初级因子。

例1 求矩阵

$$A=\begin{bmatrix}-1 & & & \\ & -2 & & \\ & & 1 & \\ & & & 2\end{bmatrix}$$

的不变因式及初级因子。

解 因为 A 的特征矩阵为

$$A(\lambda)=\lambda E-A=\begin{bmatrix}\lambda+1 & & & \\ & \lambda+2 & & \\ & & \lambda-1 & \\ & & & \lambda-2\end{bmatrix}$$

所以 $A(\lambda)$的行列式因子为：

$$D_4(\lambda)=|\lambda E-A|=(\lambda^2-1)(\lambda^2-4),$$
$$D_3(\lambda)=D_2(\lambda)=D_1(\lambda)=1;$$

不变因式为：

$$d_1(\lambda)=D_1(\lambda)=1, d_2(\lambda)=d_3(\lambda)=1,$$
$$d_4(\lambda)=D_4(\lambda)/D_3(\lambda)=(\lambda-1)(\lambda+1)(\lambda-2)(\lambda+2).$$

而次数大于零的不变因式只有 $d_4(\lambda)$，故由定义可见 A 的全部初级因子为：

$$\lambda-1, \lambda+1, \lambda-2, \lambda+2.$$

例2 设

$$A=\begin{bmatrix} a & -b_1 & & & 0 \\ & a & \ddots & & \\ & & \ddots & & -b_{n-1} \\ 0 & & & & a \end{bmatrix}_{n\times n} \quad (\text{各个 } b_i\neq 0),$$

求 A 的初级因子。

解 因为

$$A(\lambda)=\lambda E-A=\begin{bmatrix} \lambda-a & b_1 & & \\ & \lambda-a & \ddots & \\ & & \ddots & b_{n-1} \\ & & & \lambda-a \end{bmatrix}_{n\times n}$$

所以 $D_n(\lambda)=|\lambda E-A|=(\lambda-a)^n$。又因它有一个 $n-1$ 级子式

$$\begin{vmatrix} b_1 & & & 0 \\ & b_2 & & \\ & & \ddots & \\ * & & & b_{n-1} \end{vmatrix}=b_1b_2\cdots b_{n-1}\neq 0,$$

故 $D_{n-1}(\lambda)=1$，从而 $D_{n-2}(\lambda)=\cdots=D_1(\lambda)=1$。于是，不变因式为：

$$d_1(\lambda)=d_2(\lambda)=\cdots=d_{n-1}(\lambda)=1,d_n(\lambda)=(\lambda-a)^n,$$

因此初级因子只有一个：$(\lambda-a)^n$。

例 3 求下列矩阵的初级因子：

1) $A=\begin{bmatrix} 1 & 2 & 0 \\ 0 & 2 & 0 \\ -2 & -2 & -1 \end{bmatrix}$ 2) $A=\begin{bmatrix} -1 & -2 & -1 & 0 \\ 2 & -1 & 0 & -1 \\ 0 & 0 & -1 & -2 \\ 0 & 0 & 2 & -1 \end{bmatrix}$

解 1) 因

$$A(\lambda)=\lambda E-A=\begin{bmatrix} \lambda-1 & -2 & 0 \\ 0 & \lambda-2 & 0 \\ 2 & 2 & \lambda+1 \end{bmatrix}$$

故 $D_3(\lambda)=|\lambda E-A|=(\lambda-1)(\lambda+1)(\lambda-2)$。又易得 $D_2(\lambda)=1$，从

而 $D_1(\lambda)=1$. 于是，不变因式为：

$$1,1,(\lambda-1)(\lambda+1)(\lambda-2),$$

而初级因子为：$\lambda-1,\lambda+1,\lambda-2$。

2) 写出 $A(\lambda)=\lambda E-A$ 的式子，可见

$$D_4(\lambda)=|\lambda E-A|$$

$$=\begin{vmatrix}\lambda+1 & 2\\ -2 & \lambda+1\end{vmatrix}\cdot\begin{vmatrix}\lambda+1 & 2\\ -2 & \lambda+1\end{vmatrix}=[(\lambda+1)^2+4]^2。$$

又 $A(\lambda)$中有一个三阶子式为

$$\begin{vmatrix}2 & 1 & 0\\ \lambda+1 & 0 & 1\\ 0 & \lambda+1 & 2\end{vmatrix}=-4(\lambda+1),$$

但另一个三阶子式为

$$\begin{vmatrix}-2 & 0 & 1\\ 0 & \lambda+1 & 2\\ 0 & -2 & \lambda+1\end{vmatrix}=-2[(\lambda+1)^2+4]。$$

这两个三阶子式互素（不计常数因子），故知 $D_3(\lambda)=1$，从而 $D_2(\lambda)=D_1(\lambda)=1$. 因此，不变因式为：

$$1,1,1,[(\lambda+1)^2+4]^2$$
$$=[\lambda-(-1+2i)]^2\cdot[\lambda-(-1-2i)]^2,$$

所以 A 的初级因子为：

$$[\lambda-(-1+2i)]^2,[\lambda-(-1-2i)]^2,$$

亦即 $(\lambda+1-2i)^2,(\lambda+1+2i)^2$。

有了上述概念，现在来考虑矩阵 A 的标准形问题。

设矩阵 $A=(a_{ij})\in C^{n\times n}$的全部初级因子为：

$$(\lambda-\lambda_1)^{k_1},(\lambda-\lambda_2)^{k_2},\cdots,(\lambda-\lambda_s)^{k_s}.$$

在这里，$\lambda_1,\lambda_2,\cdots,\lambda_s$ 可能有相同的，指数 $k_1,k_2,\cdots,k_s$ 也可能有相同的. 对每个初级因子 $(\lambda-\lambda_i)^{k_i}$构作一个 k_i 阶矩阵（约当块）：

$$J_i = \begin{bmatrix} \lambda_i & & & & 0 \\ 1 & \lambda_i & & & \\ & 1 & \ddots & & \\ & 0 & \ddots & \ddots & \\ & & & 1 & \lambda_i \end{bmatrix} \quad (i = 1,2,\cdots,s).$$

由所有这些约当块构成的分块对角矩阵

$$J = \begin{bmatrix} J_1 & & & \\ & J_2 & & \\ & & \ddots & \\ & & & J_s \end{bmatrix}$$

称为矩阵 A 的**约当形矩阵**，或 A 的**约当标准形**。

现在我们来叙述关于约当矩阵的基本定理。

定理 1 每个 n 阶复数矩阵 A 都与一个约当形矩阵 J 相似：

$$P^{-1}AP = J;$$

除去约当块的排列次序外，约当形矩阵 J 是被矩阵 A 唯一决定的。

这个定理用线性变换的语言来说就是：

设 T 是复数域上 n 维线性空间 V 的线性变换，则在 V 中必定存在一组基，使 T 在这组基下的矩阵是约当形矩阵；除去约当块的排列次序外，这个约当形矩阵是被 T 唯一决定的。

因为对角形矩阵是约当形矩阵的特殊情形，即它是由 n 个一阶约当块构成的约当形矩阵。因此我们有：

推论 复数矩阵 A 与对角形矩阵相似的充要条件，是 A 的初级因子全为一次式。

注意，由于

$$\begin{aligned} |\lambda E - A| &= |\lambda E - J| \\ &= |\lambda E_1 - J_1| \cdot |\lambda E_2 - J_2| \cdots |\lambda E_s - J_s| \\ &= (\lambda - \lambda_1)^{k_1}(\lambda - \lambda_2)^{k_2}\cdots(\lambda - \lambda_s)^{k_s} \end{aligned}$$

所以约当形矩阵 J 的主对角线上的元素 $\lambda_1,\lambda_2,\cdots,\lambda_s$ 全为 A 的特

征值，并且$\sum_{i=1}^{s}k_i=n$。但由于$i\neq j$时可能有$\lambda_i=\lambda_j$，故λ_i不一定是A的k_i重特征根(见下面例5)，故一般由矩阵的特征多项式是不能写出矩阵的约当形矩阵的。这点常为学生所误解，请加以注意。

回头看前述三例所给矩阵的约当标准形。

例1的矩阵A的约当标准形是

$$J=\begin{bmatrix}1 & & & \\ & -1 & & \\ & & 2 & \\ & & & -2\end{bmatrix}$$

例2的约当标准形为

$$J=\begin{bmatrix}a & & & & \\ 1 & a & & & \\ & 1 & \ddots & & \\ & & \ddots & a & \\ & & & 1 & a\end{bmatrix}_{n\times n}$$

例3的两个矩阵的约当标准形分别是：

1) $J=\begin{bmatrix}1 & & \\ & -1 & \\ & & 2\end{bmatrix}$；

2) $J=\begin{bmatrix}-1+2i & & & \\ 1 & -1+2i & & \\ & & -1-2i & \\ & & 1 & -1-2i\end{bmatrix}$

下面再举几个例子。

例4 求矩阵

$$A=\begin{bmatrix}2 & -1 & -1\\ 2 & -1 & -2\\ -1 & 1 & 2\end{bmatrix}$$

的约当标准形及所用的矩阵P。

解 因为

$$\lambda E - A = \begin{bmatrix} \lambda - 2 & 1 & 1 \\ -2 & \lambda + 1 & 2 \\ 1 & -1 & \lambda - 2 \end{bmatrix}.$$

的初级因子为$(\lambda-1)$,$(\lambda-1)^2$,故A的约当标准形为:

$$J = \begin{bmatrix} 1 & & \\ & 1 & \\ & 1 & 1 \end{bmatrix}$$

再设$P=(X_1,X_2,X_3)$,由

$$P^{-1}AP = J$$

得

$$A(X_1,X_2,X_3) = (X_1,X_2,X_3)J,$$

于是有

$$(AX_1,AX_2,AX_3) = (X_1,X_2 + X_3,X_3),$$

即得:

$$(E - A)X_1 = \theta, \tag{1}$$

$$(E - A)X_2 = - X_3, \tag{2}$$

$$(E - A)X_3 = \theta. \tag{3}$$

解方程(1)得基础解系为

$$e_1 = (1,1,0)^T,\ e_2 = (1,0,1)^T.$$

我们选取$X_1=(1,1,0)^T$。又由于方程(3)与(1)是一样的,所以(3)的任一解具有形式

$$X_3 = c_1e_1 + c_2e_2 = (c_1 + c_2,c_1,c_2)^T.$$

为使方程(2)有解,可选择c_1,c_2的值使下面两矩阵的秩相等:

$$E - A = \begin{bmatrix} -1 & 1 & 1 \\ -2 & 2 & 2 \\ 1 & -1 & -1 \end{bmatrix}, \quad \begin{bmatrix} -1 & 1 & 1 & c_1 + c_2 \\ -2 & 2 & 2 & c_1 \\ 1 & -1 & -1 & c_2 \end{bmatrix}.$$

这样可得$c_1=2,c_2=-1$。因而

$$X_3 = (1,2,-1)^T.$$

将所求得的 X_3 代入方程(2),并解之便得

$$X_2 = (1,1,1)^T.$$

易证 X_1,X_2,X_3 线性无关,故取 $P=(X_1,X_2,X_3)$,即

$$P = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 1 \\ 0 & -1 & 1 \end{bmatrix}$$

便有 $P^{-1}AP=J$.

例5 求矩阵

$$A = \begin{bmatrix} 4 & 6 & 0 \\ -3 & -5 & 0 \\ -3 & -6 & 1 \end{bmatrix}$$

的特征多项式、初级因子及约当标准形.

解 易得特征多项式为

$$f(\lambda) = |\lambda E - A| = (\lambda-1)^2 \cdot (\lambda+2),$$

并且可以求得不变因式为:

$$d_1(\lambda) = 1, d_2(\lambda) = \lambda - 1, d_3(\lambda) = (\lambda-1)(\lambda+2).$$

故初级因子为:

$$\lambda-1, \lambda-1, \lambda+2.$$

因此约当标准形为对角形矩阵:

$$J = \begin{bmatrix} 1 & & \\ & 1 & \\ & & -2 \end{bmatrix}$$

例6 求线性微分方程组

$$\frac{dx_1}{dt} = -x_1 + x_2,$$

$$\frac{dx_2}{dt} = -4x_1 + 3x_2,$$

$$\frac{\mathrm{d}x_3}{\mathrm{d}t}=x_1+2x_3.$$

的通解。这里 x_1,x_2,x_3 都是 t 的未知函数。

解 这方程组写成矩阵形式为

$$\frac{\mathrm{d}X}{\mathrm{d}t}=AX.$$

其中

$$A=\begin{bmatrix}-1 & 1 & 0\\ -4 & 3 & 0\\ 1 & 0 & 2\end{bmatrix}\qquad X=(x_1,x_2,x_3)^T.$$

(向量的导数定义为每个分量的导数,见下章)

我们不难求得 A 的初级因子:

$$\lambda-2,(\lambda-1)^2.$$

于是有可逆矩阵 $P=(X_1,X_2,X_3)$,使得

$$P^{-1}AP=J=\begin{bmatrix}2 & 0 & 0\\ 0 & 1 & 0\\ 0 & 1 & 1\end{bmatrix}$$

由 $AP=PJ$,即有

$$(AX_1,AX_2,AX_3)=(2X_1,X_2+X_3,X_3),$$

所以有

$$\begin{cases}AX_1=2X_1\\ AX_2=X_2+X_3\\ AX_3=X_3\end{cases}$$

求得:

$$X_1=(0,0,1)^T,X_2=(0,1,-1)^T,X_3=(1,2,-1)^T.$$

所以有:

$$P=\begin{bmatrix}0 & 0 & 1\\ 0 & 1 & 2\\ 1 & -1 & -1\end{bmatrix}$$

作满秩线性变换

$$X=PY,\quad 其中\ Y=(y_1,y_2,y_3)^T.$$

则有

$$APY=AX=\frac{\mathrm{d}X}{\mathrm{d}t}=P\frac{\mathrm{d}Y}{\mathrm{d}t}.$$

所以

$$\frac{\mathrm{d}Y}{\mathrm{d}t}=P^{-1}APY=\begin{bmatrix}2 & 0 & 0\\ 0 & 1 & 0\\ 0 & 1 & 1\end{bmatrix}Y$$

即

$$\begin{cases}\dfrac{\mathrm{d}y_1}{\mathrm{d}t}=2y_1\\ \dfrac{\mathrm{d}y_2}{\mathrm{d}t}=y_2\\ \dfrac{\mathrm{d}y_3}{\mathrm{d}t}=y_2+y_3\end{cases}$$

积分前两个方程可得

$$y_1=k_1\mathrm{e}^{2t},\ y_2=k_2\mathrm{e}^{t}.$$

将求得的 y_2 代入第三个方程并积分它便得

$$y_3=(k_2t+k_3)\mathrm{e}^{t},$$

故微分方程组的通解为：

$$X=PY=\begin{bmatrix}0 & 0 & 1\\ 0 & 1 & 2\\ 1 & -1 & -1\end{bmatrix}\begin{bmatrix}k_1\mathrm{e}^{2t}\\ k_2\mathrm{e}^{t}\\ (k_2t+k_3)\mathrm{e}^{t}\end{bmatrix}$$

$$=\begin{bmatrix}(k_2t+k_3)e^t\\(2k_2t+k_2+2k_3)e^t\\k_1e^{2t}-(k_2t+k_2+k_3)e^t\end{bmatrix}$$

这里 k_1,k_2,k_3 是任意常数。

例 7 利用约当形矩阵证明：若 n 阶矩阵 A 的特征值为 $\lambda_1,\lambda_2,\cdots,\lambda_n$，则矩阵 A^m 的特征值为 $\lambda_1^m,\lambda_2^m,\cdots,\lambda_n^m$。

证明 设 A 的约当形矩阵为

$$J=\begin{bmatrix}J_1&&&\\&J_2&&\\&&\ddots&\\&&&J_s\end{bmatrix}\qquad \text{其中 } J_i=\begin{bmatrix}\lambda_i&&&&\\1&\lambda_i&&&\\&1&\ddots&&\\&&\ddots&\ddots&\\&&&1&\lambda_i\end{bmatrix}$$

因 $J=P^{-1}AP$，故

$$J^m=P^{-1}A^mP\ .$$

但是我们有

$$J^m=\begin{bmatrix}J_1^m&&&\\&J_2^m&&\\&&\ddots&\\&&&J_s^m\end{bmatrix}\qquad J_i^m=\begin{bmatrix}\lambda_i^m&&&\\ *&\lambda_i^m&&\\ \vdots&\ddots&\ddots&\\ *&\cdots&*&\lambda_i^m\end{bmatrix}$$

显然 J^m 的特征值就是 J 的特征值的 m 次幂，而相似矩阵有相同的特征值，故 A^m 的特征值就是 J^m 的特征值，即 A（或 J）的特征值的 m 次幂。证毕。

§3 哈密顿-开莱定理及矩阵的最小多项式

在第一节我们给出了矩阵的特征多项式，本节将进一步给出特征多项式的性质，其中最重要的就是哈密顿-开莱定理；我们还将讨论另一个重要的多项式即矩阵的最小多项式，及其与特征多项式的关系。本节所得到的结果有重要的理论及应用价值。

定理 1 每个 n 阶矩阵 A 都是它的特征多项式的根，即

$$A^n + a_1A^{n-1} + \cdots + a_{n-1}A + a_nE = 0(\text{矩阵}). \quad (1)$$

这个定理称为哈密顿-开莱(Hamilton—Cayley)定理。因为 $f(\lambda)=|\lambda E-A|=\lambda^n+a_1\lambda^{n-1}+\cdots+a_{n-1}\lambda+a_n$，所以(1)式常写成 $f(A)=0$(矩阵)。

证明 设 $B(\lambda)$ 为 $\lambda E-A$ 的伴随矩阵，则

$$B(\lambda)(\lambda E - A) = |\lambda E - A|E = f(\lambda)E. \quad (2)$$

由于矩阵 $B(\lambda)$ 的元素都是行列式 $|\lambda E-A|$ 中的元素的代数余子式，因而都是 λ 的多项式，其次数都不超过 $n-1$。故由矩阵运算性质，$B(\lambda)$ 可以写成如下形式：

$$B(\lambda) = \lambda^{n-1}B_0 + \lambda^{n-2}B_1 + \cdots + B_{n-1},$$

这里各个 B_i 均为 n 阶数字矩阵。因此有

$$B(\lambda)(\lambda E - A) = \lambda^nB_0 + \lambda^{n-1}(B_1 - B_0A) + \cdots + \lambda(B_{n-1} - B_{n-2}A) - B_{n-1}A. \quad (3)$$

另方面又显然有

$$f(\lambda)E = \lambda^nE + a_1\lambda^{n-1}E + \cdots + a_{n-1}\lambda E + a_nE. \quad (4)$$

由等式(2)、(3)、(4)即得：

$$\begin{cases} B_0 = E \\ B_1 - B_0A = a_1E \\ \cdots\cdots\cdots\cdots \\ B_{n-1} - B_{n-2}A = a_{n-1}E \\ - B_{n-1}A = a_nE \end{cases} \quad (5)$$

以 $A^n, A^{n-1}, \cdots, A, E$ 依次右乘(5)的第一式，第二式，…，第 $n+1$ 式，并将它们加起来，则左边变成零矩阵，而右边即为 $f(A)$，故有 $f(A)=0$(矩阵)。证毕。

注意，上述证明 $f(A)=0$(矩阵)的过程有时会被人认为麻烦，而认为：由于 $f(\lambda)=|\lambda E-A|$，所以"显然"有 $f(A)=|AE-A|=0$(矩阵)。这样做对吗？

例 1 设

$$A=\begin{bmatrix}1 & 0 & 2\\ 0 & -1 & 1\\ 0 & 1 & 0\end{bmatrix}$$

试计算 $\varphi(A)=2A^8-3A^5+A^4+A^2-4E$。

解 因 A 的特征多项式为

$$f(\lambda)=|\lambda E-A|=\lambda^3-2\lambda+1\ .$$

我们再取多项式

$$\varphi(\lambda)=2\lambda^8-3\lambda^5+\lambda^4+\lambda^2-4\ .$$

以 $f(\lambda)$ 去除 $\varphi(\lambda)$ 可得

$$\varphi(\lambda)=(2\lambda^5+4\lambda^3-5\lambda^2+9\lambda-14)f(\lambda)+r(\lambda),$$

这里余式 $r(\lambda)=24\lambda^2-37\lambda+10$。

由哈密顿—开莱定理，$f(A)=0$（矩阵），所以

$$\varphi(A)=r(A)=24A^2-37A+10E$$

$$=\begin{bmatrix}-3 & 48 & -26\\ 0 & 95 & -61\\ 0 & -61 & 34\end{bmatrix}$$

这例子说明，一个 n 阶矩阵的多项式，如其次数高于 n 次，则应用定理 1 可将它化为次数 $<n$ 的多项式来计算。

一般地说，若 $\varphi(\lambda)$ 是个多项式，A 是个方阵，如果有 $\varphi(A)=0$（矩阵），则称 $\varphi(\lambda)$ 是矩阵 A 的*零化多项式*。显然，每个方阵都有零化多项式，因为它的特征多项式就是一个，但并不唯一。下述零化多项式亦很有用。

定义 1 设 A 是 n 阶矩阵，则 A 的首项系数为 1 的次数最小的零化多项式 $m(\lambda)$，称为 A 的*最小多项式*。

首项（最高次项）系数为 1 的多项式，以后简称*首一多项式*。

定理 2 矩阵 A 的任何零化多项式都被其最小多项式所整除。

证明 设$\varphi(\lambda)$是A的任一零化多项式，又$m(\lambda)$是A的最小多项式。以$m(\lambda)$除$\varphi(\lambda)$即得

$$\varphi(\lambda) = q(\lambda)m(\lambda) + r(\lambda),$$

这里$r(\lambda)$如不为零时则其次数小于$m(\lambda)$的次数。于是有

$$\varphi(A) = q(A)m(A) + r(A)。$$

因$\varphi(A)=m(A)=0$(矩阵)，所以有$r(A)=0$(矩阵)，即$r(\lambda)$也是A的零化多项式。如果$r(\lambda)\neq 0$，则$r(\lambda)$的次数$<m(\lambda)$的次数，这与$m(\lambda)$为最小多项式矛盾。所以，只能有$r(\lambda)\equiv 0$，故$m(\lambda)|\varphi(\lambda)$。证毕。

定理 3 矩阵A的最小多项式是唯一的。

证明 若$m(\lambda)$与$n(\lambda)$均为A的最小多项式。那末每一个都可被另一个所整除，因此两者只有常数因子的差别。这常数因子必定等于1，因为两者都是首一多项式。故$m(\lambda)=n(\lambda)$。

定理 4 矩阵A的最小多项式的根必定是A的特征根；反之，A的特征根也必定是A的最小多项式的根。

证明 因A的特征多项式$f(\lambda)=|\lambda E-A|$是A的零化多项式，故由定理2，$f(\lambda)$可被A的最小多项式$m(\lambda)$所整除，即$m(\lambda)$是$f(\lambda)$的因式，所以$m(\lambda)$的根都是$f(\lambda)$的根。

反之，若λ_0是A的一个特征根，且

$$AX = \lambda_0 X, \quad (X \neq \theta)。$$

又设A的最小多项式

$$m(\lambda) = \lambda^k + \beta_1\lambda^{k-1} + \cdots + \beta_{k-1}\lambda + \beta_k.$$

则

$$\begin{aligned} m(A)X &= A^kX + \beta_1A^{k-1}X + \cdots + \beta_{k-1}AX + \beta_kX \\ &= \lambda_0^kX + \beta_1\lambda_0^{k-1}X + \cdots + \beta_{k-1}\lambda_0X + \beta_kX \\ &= m(\lambda_0)X . \end{aligned}$$

由于$m(A)=0$(矩阵)，又$X\neq\theta$，所以$m(\lambda_0)=0$，亦即λ_0是$m(\lambda)$的根。证毕。

这定理反映了特征多项式与最小多项式之间的重要关系。由

此可得到求最小多项式的一个方法。

设矩阵 $A\in C^{n\times n}$的所有不同的特征值为 $\lambda_1,\lambda_2,\cdots,\lambda_s$，又 A 的特征多项式为

$$f(\lambda)=|\lambda E-A|=(\lambda-\lambda_1)^{k_1}(\lambda-\lambda_2)^{k_2}\cdots(\lambda-\lambda_s)^{k_s},$$

则 A 的最小多项式必具有如下形式：

$$m(\lambda)=(\lambda-\lambda_1)^{n_1}(\lambda-\lambda_2)^{n_2}\cdots(\lambda-\lambda_s)^{n_s},$$

这里每个 $n_i\leqslant k_i, i=1,2,\cdots,s$。

例 2 求矩阵

$$A=\begin{bmatrix}3 & -3 & 2\\ -1 & 5 & -2\\ -1 & 3 & 0\end{bmatrix}$$

的最小多项式 $m(\lambda)$。

解 A 的特征多项式为

$$f(\lambda)=|\lambda E-A|=(\lambda-2)^2(\lambda-4),$$

故 A 的最小多项式只能是

$$m(\lambda)=(\lambda-2)(\lambda-4),\text{或 } m(\lambda)=f(\lambda).$$

但由

$$m(A)=(A-2E)(A-4E)=0(\text{矩阵})(\text{直接计算可得}),$$

便知 A 的最小多项式应为 $m(\lambda)=(\lambda-2)(\lambda-4)$而不是 $f(\lambda)$。

下面的定理 5 反映了矩阵的最小多项式与特征矩阵的行列式因子及不变因式的关系，并提供了求最小多项式的另一方法。

定理 5 设 A 是 n 阶矩阵，$D_{n-1}(\lambda)$是特征矩阵 $\lambda E-A$ 的 $n-1$ 阶行列式因子，则 A 的最小多项式

$$m(\lambda)=\frac{|\lambda E-A|}{D_{n-1}(\lambda)}=\frac{D_n(\lambda)}{D_{n-1}(\lambda)}=d_n(\lambda),$$

这里 $d_n(\lambda)$是 $\lambda E-A$ 的第 n 个不变因式。(证略)。

§4 多项式矩阵与史密斯标准形

若矩阵 $A(\lambda)=(a_{ij}(\lambda))_{m\times n}$的元素 $a_{ij}(\lambda)$都是文字 λ 的多项式(系数属于某一数域 P),则 $A(\lambda)$称为 λ-矩阵,或多项式矩阵。在§2 中我们曾讨论过它的一个很特别的情形,即 $A(\lambda)=\lambda E-A$ 为数字矩阵 A 的特征矩阵的情形。在这一节,我们将介绍一般的多项式矩阵的基本理论。这个理论在线性控制系统理论中有着重要的应用。

作为多项式矩阵的一种推广就是有理分式矩阵。以传递函数描述方法为基础的线性系统复频率域理论,研究对象为线性定常系统,基本的系统模型是传递函数矩阵(有理分式矩阵)的矩阵分式描述,主要数学工具是多项式矩阵理论。采用矩阵分式描述,使得有可能在多项式矩阵理论的基础上,建立起分析与综合线性系统的一整套比较简便和实用的理论和方法。此外,在线性系统理论中,也常使用多项式矩阵描述,它与其他形式的描述(状态空间描述,矩阵分式描述等)之间,不仅在形式上而且在有关的系统结构特性上,都可找到确定的内在联系及等价关系。

若多项式矩阵 $A(\lambda)$至少有一个 $r(\geqslant 1)$阶子式不是零多项式,而一切 $r+1$ 阶子式(如有的话)都是零多项式,则称 $A(\lambda)$的秩是 r。零矩阵的秩定义为零。

例如,数字矩阵 $A=(a_{ij})_{n\times n}$的特征矩阵 $A(\lambda)=\lambda E-A$ 的秩是 n,因为 $f(\lambda)=|\lambda E-A|$不是零多项式(至多有 n 个根)。

一般,若 n 阶多项式矩阵 $A(\lambda)$的行列式$|A(\lambda)|$不等于零多项式,则称 $A(\lambda)$是满秩的(秩$=n$),或非奇异的。如果存在多项式矩阵 $B(\lambda)$,使得

$$A(\lambda)B(\lambda)=B(\lambda)A(\lambda)=E, \qquad (1)$$

则称 $A(\lambda)$是可逆的,或称 $A(\lambda)$是单模矩阵,这里 E 为 n 阶单位矩

阵。

与数字矩阵的证法一样，满足(1)式的 $B(\lambda)$是唯一的，并记为 $A^{-1}(\lambda)$，它称为 $A(\lambda)$的逆矩阵。

定理 1 n 阶多项式矩阵 $A(\lambda)$可逆的充要条件，是 $A(\lambda)$的行列式等于非零常数：$|A(\lambda)|=c\neq0$。

证明 设 $A(\lambda)$可逆，则有多项式矩阵 $B(\lambda)$，使得(1)式成立，从而有

$$|A(\lambda)|\cdot|B(\lambda)| = |E| = 1 .$$

故$|A(\lambda)|$与$|B(\lambda)|$只能是零次多项式，且不等于零(数)，所以当 $A(\lambda)$可逆时，$|A(\lambda)|$必定等于某个非零常数 c。

反过来，若$|A(\lambda)|=c\neq0$，则易知 $A(\lambda)$可逆，且其逆矩阵为

$$A^{-1}(\lambda) = \frac{1}{c}A^{*}(\lambda),$$

这里 $A^{*}(\lambda)$是 $A(\lambda)$的伴随矩阵。证毕。

由这个定理可见，在多项式矩阵中，满秩矩阵未必是可逆的，如上面提到的特征矩阵 $\lambda E-A$ 虽然满秩，但不可逆。可逆(单模)的要求更高了。这与数字矩阵是不同的。

例 1 多项式矩阵

$$A(\lambda) = \begin{pmatrix} \lambda+1 & \lambda+3 \\ \lambda^2+3\lambda & \lambda^2+5\lambda+4 \end{pmatrix}$$

$$B(\lambda) = \begin{pmatrix} \lambda+1 & \lambda+3 \\ \lambda^2+3\lambda+2 & \lambda^2+5\lambda+6 \end{pmatrix}$$

中，$A(\lambda)$是可逆的，而 $B(\lambda)$是不可逆的，因为

$$|A(\lambda)| = 4, \quad |B(\lambda)| = 0 .$$

定义 1 下列运算，称为多项式矩阵 $A(\lambda)$的初等变换：

(1) 互换 $A(\lambda)$的任意两行(列)；

(2) 以非零的数 $c(\in P)$乘 $A(\lambda)$的某一行(列)；

(3) 以多项式 $\varphi(\lambda)$乘 $A(\lambda)$的某一行(列)并加到另一行(列)

上。

由单位矩阵 E 经过一次上述初等变换得到的矩阵称为*初等矩阵*。容易验证，初等矩阵都是可逆的，即它们都是单模矩阵，事实上，若用 $E(i,j)$，$E(i(c))$，$E(i,\varphi(\lambda)j)$分别表示由单位矩阵 E 互换 i,j 两行(列)；第 i 行(列)乘以非零常数 c；第 j 行(列)乘以多项式 $\varphi(\lambda)$并加到第 i 行(列)上所得到的初等矩阵，则有

$$|E(i,j)| = -1, |E(i(c))| = c \neq 0, |E(i,\varphi(\lambda)j| = 1 .$$

且可以验证

$$E(i,j)^{-1} = E(i,j); E(i(c))^{-1} = E(i(c^{-1}));$$

$$E(i,\varphi(\lambda)j)^{-1} = E(i,-\varphi(\lambda)j) .$$

与数字矩阵的情形一样，可以证明：对一个多项式矩阵 $A(\lambda)$ 进行一次初等行(列)变换，相当于用一个相应的初等矩阵左(右)乘矩阵 $A(\lambda)$。

定义 2 多项式矩阵 $A(\lambda)$称为与多项式矩阵 $B(\lambda)$等价，如果经过有限次初等变换能把 $A(\lambda)$化为 $B(\lambda)$。

当 $A(\lambda)$与 $B(\lambda)$等价时，就记为 $A(\lambda) \cong B(\lambda)$。

容易验证，多项式矩阵的这一等价定义，具有下述性质：

1) 自反性 $A(\lambda) \cong A(\lambda)$。

2) 对称性 $A(\lambda) \cong B(\lambda) \Rightarrow B(\lambda) \cong A(\lambda)$。

3) 传递性 若 $A(\lambda) \cong B(\lambda)$，且 $B(\lambda) \cong C(\lambda)$，则 $A(\lambda) \cong C(\lambda)$。

简言之，$A(\lambda) \cong B(\lambda)$是一种等价关系。

不难证明 $A(\lambda) \cong B(\lambda)$的充分必要条件，是存在初等矩阵 $P_1, \cdots, P_s, Q_1, \cdots, Q_t$，使得

$$B(\lambda) = P_1 \cdots P_s A(\lambda) Q_1 \cdots Q_t = P(\lambda) A(\lambda) Q(\lambda),$$

这里 $P(t)$，$Q(t)$都是单模矩阵。

在多项式矩阵的应用中，有多种标准形在不同场合里被使用着，在这里我们只介绍其中最基本的一种，即史密斯(Smith)标准

形。我们用定理形式给出这一标准形，由于它的重要性，将给出详细的证明，但要先证明一个引理。

引理 若多项式矩阵 $A(\lambda)=(a_{ij}(\lambda))_{m\times n}$的左上角元素 $a_{11}(\lambda)\neq 0$，并且 $A(\lambda)$中至少有一个元素不能被 $a_{11}(\lambda)$所整除，则必可找到一个与 $A(\lambda)$等价的多项式矩阵 $B(\lambda)$，其左上角元素 $b_{11}(\lambda)$也不等于零，且 $b_{11}(\lambda)$的次数低于 $a_{11}(\lambda)$的次数。

证明 我们可分三种情况来讨论：

1) 若 $A(\lambda)$的第一列中有某个元素 $a_{i1}(\lambda)$不能被 $a_{11}(\lambda)$整除，则用 $a_{11}(\lambda)$去除 $a_{i1}(\lambda)$可得

$$a_{i1}(\lambda)=q(\lambda)a_{11}(\lambda)+r(\lambda),$$

且余式 $r(\lambda)(\neq 0)$的次数低于 $a_{11}(\lambda)$的次数。此时则有

$$A(\lambda)=\begin{bmatrix} a_{11}(\lambda) & a_{12}(\lambda) & \cdots & a_{1n}(\lambda) \\ \cdots & \cdots & \cdots & \cdots \\ a_{i1}(\lambda) & a_{i2}(\lambda) & \cdots & a_{in}(\lambda) \\ \cdots & \cdots & \cdots & \cdots \\ a_{m1}(\lambda) & a_{m2}(\lambda) & \cdots & a_{mn}(\lambda) \end{bmatrix}$$

$$\xrightarrow[\substack{((i)\text{ 行减去}(1)\\ \text{行的 } q(\lambda)\text{ 倍})}]{[(i)-q(\lambda)(1)]}\begin{bmatrix} a_{11}(\lambda) & a_{12}(\lambda) & \cdots\cdots & a_{1n}(\lambda) \\ \cdots & \cdots & \cdots & \cdots \\ r(\lambda) & a_{i2}(\lambda)-q(\lambda)a_{12}(\lambda) & \cdots\cdots & \\ \cdots & \cdots & \cdots & \cdots \\ a_{m1}(\lambda) & a_{m2}(\lambda) & \cdots\cdots & a_{mn}(\lambda) \end{bmatrix}$$

$$\xrightarrow[(\text{交换}(1)\text{ 行与}(i)\text{ 行})]{[(1),(i)]}\begin{bmatrix} r(\lambda) & a_{i2}(\lambda)-q(\lambda)a_{12}(\lambda) & \cdots\cdots & \\ \cdots & \cdots & \cdots & \cdots \\ a_{11}(\lambda) & a_{12}(\lambda) & \cdots\cdots & a_{1n}(\lambda) \\ \cdots & \cdots & \cdots & \cdots \\ a_{m1}(\lambda) & a_{m2}(\lambda) & \cdots\cdots & a_{mn}(\lambda) \end{bmatrix}$$

若上面最后得到的矩阵记为 $B(\lambda)$，则 $B(\lambda)$已达到要求；

2) 若 $A(\lambda)$的第一行中有某个元素 $a_{1j}(\lambda)$不能被 $a_{11}(\lambda)$整除，

则证法与 1)类似；

3) 若 $A(\lambda)$的第一行与第一列的各个元素均可被 $a_{11}(\lambda)$整除，但 $A(\lambda)$中至少有某个元素 $a_{ij}(\lambda)(i,j>1)$不能被 $a_{11}(\lambda)$整除. 此时可设

$$a_{i1}(\lambda)=\varphi(\lambda)a_{11}(\lambda),$$

则有

$$A(\lambda)\xrightarrow{[(i)-\varphi(\lambda)(1)]}\begin{bmatrix} a_{11}(\lambda) & \cdots & a_{1j}(\lambda) & \cdots & a_{1n}(\lambda) \\ \cdots\cdots & \cdots\cdots & \cdots\cdots & \cdots\cdots & \cdots\cdots \\ 0 & \cdots & a_{ij}(\lambda)-\varphi(\lambda)a_{1j}(\lambda) & & \cdots \\ \cdots\cdots & \cdots\cdots & \cdots\cdots & \cdots\cdots & \cdots\cdots \\ a_{m1}(\lambda) & \cdots & a_{mj}(\lambda) & \cdots & a_{mn}(\lambda) \end{bmatrix}$$

$$\xrightarrow[\text{(第 } i \text{ 行加到第 1 行)}]{[(1)+(i)]}\begin{bmatrix} a_{11}(\lambda) & \cdots & a_{ij}(\lambda)+(1-\varphi(\lambda))a_{1j}(\lambda) & & \cdots \\ \cdots\cdots & \cdots\cdots & \cdots\cdots & \cdots\cdots & \cdots\cdots \\ 0 & \cdots & a_{ij}(\lambda)-\varphi(\lambda)a_{1j}(\lambda) & & \cdots \\ \cdots\cdots & \cdots\cdots & \cdots\cdots & \cdots\cdots & \cdots\cdots \\ a_{m1}(\lambda) & \cdots & a_{mj}(\lambda) & \cdots & a_{mn}(\lambda) \end{bmatrix}$$

$$=A_1(\lambda).$$

则 $A_1(\lambda)$的第一行中已至少有一个元素

$$a_{ij}(\lambda)+(1-\varphi(\lambda))a_{1j}(\lambda)=f(\lambda)$$

不能被左上角元素 $a_{11}(\lambda)$所整除(因为 $a_{11}(\lambda)\mid a_{1j}(\lambda)$，但 $a_{11}(\lambda)\nmid a_{ij}(\lambda)$. 由此易推知 $a_{11}(\lambda)\nmid f(\lambda)$)。因此情形 3)就归结为已证明了的情形 2)。证毕.

定理 2 任一非零的多项式矩阵 $A(\lambda)=(a_{ij}(\lambda))_{m\times n}$，都等价于一个如下形式的标准对角形：

$$A(\lambda) \cong J(\lambda) = \begin{bmatrix} d_1(\lambda) & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & d_2(\lambda) & \cdots & 0 & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & d_r(\lambda) & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \end{bmatrix}$$

这里 $r(\geqslant 1)$是 $A(\lambda)$的秩,$d_i(\lambda)$ $(i=1,2,\cdots,r)$是首项系数为 1 的多项式,且 $d_i(\lambda)|d_{i+1}(\lambda);i=1,2,\cdots,r-1$。

$J(\lambda)$称为 $A(\lambda)$的史密斯(Smith)标准形。

证明 因 $A(\lambda)\neq 0$(矩阵),故至少有一个非零元素,不妨设 $a_{11}(\lambda)\neq 0$(否则经行、列调动可使左上角元素不为零)。如果 $a_{11}(\lambda)$不能整除其余的各个 $a_{ij}(\lambda)$,则由引理,可用初等变换把 $A(\lambda)$化为某个多项式矩阵 $B_1(\lambda)$,并使得 $B_1(\lambda)$的左上角元素 $b_1(\lambda)\neq 0$、而且 $b_1(\lambda)$的次数低于 $a_{11}(\lambda)$的次数。若 $b_1(\lambda)$还不能整除 $B_1(\lambda)$的其他所有元素,则再次应用引理,又可找到一个与 $B_1(\lambda)$等价的矩阵 $B_2(\lambda)$,其左上角元素 $b_2(\lambda)\neq 0$,且次数低于 $b_1(\lambda)$的次数。如此下去,将会得到一系列彼此等价的多项式矩阵:

$$A(\lambda),B_1(\lambda),B_2(\lambda),\cdots.$$

这些矩阵的左上角元素均不等于零,且次数越来越低。但这些次数都是非负整数,不能无止境地降低,故在有限步以后,将会终止于一个多项式矩阵 $B_s(\lambda)$,其左上角元素 $b_s(\lambda)\neq 0$,且 $b_s(\lambda)$能整除 $B_s(\lambda)$的全部元素 $b_{ij}(\lambda)$。不妨设:

$$b_{ij}(\lambda) = q_{ij}(\lambda)b_s(\lambda)。$$

对矩阵

$$B_s(\lambda) = \begin{bmatrix} b_s(\lambda)\cdots b_{1j}(\lambda)\cdots\cdots \\ b_{i1}(\lambda)\cdots b_{ij}(\lambda)\cdots\cdots \\ \cdots\cdots\cdots\cdots\cdots\cdots \end{bmatrix}$$

的第 2 行、第 3 行、…、第 m 行依次进行适当的初等变换，然后再对第 2 列，第 3 列，…，第 n 列进行适当的初等变换，显然可把 $B_s(\lambda)$ 化为下面的形状

$$\begin{bmatrix} b_s(\lambda) & 0 & \cdots & 0 \\ 0 & & & \\ \vdots & & B_{s+1}(\lambda) & \\ 0 & & & \end{bmatrix}$$

由上面的做法过程可知 $B_{s+1}(\lambda)$ 的所有元素均可被 $b_s(\lambda)$ 所整除。对上面这个矩阵再施行一次初等变换，还可以把 $b_s(\lambda)$ 化为首一多项式 $d_1(\lambda)$。如果 $B_{s+1}(\lambda)\neq 0$（矩阵），则对 $B_{s+1}(\lambda)$ 重复上述过程，于是又得：

$$\begin{bmatrix} d_1(\lambda) & 0 & 0\cdots 0 \\ 0 & d_2(\lambda) & 0\cdots 0 \\ \vdots & 0 & \\ & \vdots & B_{s+2}(\lambda) \\ 0 & 0 & \end{bmatrix}$$

这里 $d_1(\lambda)$，$d_2(\lambda)$ 均为首一多项式，且 $d_1(\lambda)\mid d_2(\lambda)$，而 $d_2(\lambda)$ 整除 $B_{s+2}(\lambda)$ 的各个元素。如此做下去，则 $A(\lambda)$ 最后就化成了所要求的标准形。证毕。

例 2 求多项式矩阵

$$A(\lambda)=\begin{bmatrix} 0 & \lambda(\lambda-1) & 0 \\ \lambda & 0 & \lambda+1 \\ 0 & 0 & -\lambda+2 \end{bmatrix}$$

的史密斯标准形。

解 因 $A(\lambda)$ 的左上角元素为零，故先互换第一，第二行，化 $A(\lambda)$ 为

$$A_1(\lambda)=\begin{bmatrix} \lambda & 0 & \lambda+1 \\ 0 & \lambda(\lambda-1) & 0 \\ 0 & 0 & -\lambda+2 \end{bmatrix}$$

由于 $A_1(\lambda)$的左上角元素 λ 不能整除其他所有元素，故先降低它的次数，为此只须从第三列减去第一列，并交换这两列的位置即得：

$$A_2(\lambda)=\begin{bmatrix}1 & 0 & \lambda\\ 0 & \lambda(\lambda-1) & 0\\ -\lambda+2 & 0 & 0\end{bmatrix}$$

其左上角元素已能整除其他各个元素，因此应把 $A_2(\lambda)$的第一行及第一列除左上角元素外的非零元素都化为零。为此只须从第三列减去第一列的 λ 倍，然后再从第三行减去第一行的$(-\lambda+2)$倍，便得

$$A_3(\lambda)=\begin{bmatrix}1 & 0 & 0\\ 0 & \lambda(\lambda-1) & 0\\ 0 & 0 & \lambda(\lambda-2)\end{bmatrix}$$

又在 $A_3(\lambda)$中，由于 $\lambda(\lambda-1)$不能整除 $\lambda(\lambda-2)$，所以可把第三行加到第二行，然后再将第三列减去第二列，便得

$$A_4(\lambda)=\begin{bmatrix}1 & 0 & 0\\ 0 & \lambda(\lambda-1) & -\lambda\\ 0 & 0 & \lambda(\lambda-2)\end{bmatrix}$$

交换第二、第三列即得：

$$A_5(\lambda)=\begin{bmatrix}1 & 0 & 0\\ 0 & -\lambda & \lambda(\lambda-1)\\ 0 & \lambda(\lambda-2) & 0\end{bmatrix}$$

以$(\lambda-1)$乘第二列并加到第三列上，再以$(\lambda-2)$乘第二行并加到第三行上，便得

$$A_6(\lambda)=\begin{bmatrix}1 & 0 & 0\\ 0 & -\lambda & 0\\ 0 & 0 & \lambda(\lambda-1)(\lambda-2)\end{bmatrix}$$

$$\cong \begin{bmatrix} 1 & 0 & 0 \\ 0 & \lambda & 0 \\ 0 & 0 & \lambda(\lambda-1)(\lambda-2) \end{bmatrix}$$

这便是所求的标准形。而且有

$$d_1(\lambda)=1, d_2(\lambda)=\lambda, d_3(\lambda)=\lambda(\lambda-1)(\lambda-2).$$

例 3 化多项式矩阵

$$A(\lambda)=\begin{bmatrix} 1-\lambda & 2\lambda-1 & \lambda \\ \lambda & \lambda^2 & -\lambda \\ 1+\lambda^2 & \lambda^2+\lambda-1 & -\lambda^2 \end{bmatrix}$$

为史密斯标准形。

解 $A(\lambda)\xrightarrow{\text{3 列加到 1 列}}\begin{bmatrix} 1 & 2\lambda-1 & \lambda \\ 0 & \lambda^2 & -\lambda \\ 1 & \lambda^2+\lambda-1 & -\lambda^2 \end{bmatrix}$

$$\xrightarrow{\text{3 行减 1 行}}\begin{bmatrix} 1 & 2\lambda-1 & \lambda \\ 0 & \lambda^2 & -\lambda \\ 0 & \lambda^2-\lambda & -\lambda^2-\lambda \end{bmatrix}$$

$$\xrightarrow[\substack{[2,-(2\lambda-1)1] \\ [3,(-\lambda)1]}]{}\begin{bmatrix} 1 & 0 & 0 \\ 0 & \lambda^2 & -\lambda \\ 0 & \lambda^2-\lambda & -\lambda^2-\lambda \end{bmatrix}$$

$$\xrightarrow{[2,(-1)3]}\begin{bmatrix} 1 & 0 & 0 \\ 0 & \lambda & \lambda^2 \\ 0 & \lambda^2-\lambda & -\lambda^2-\lambda \end{bmatrix}$$

$$\xrightarrow{[3,(1-\lambda)2]}\begin{bmatrix} 1 & 0 & 0 \\ 0 & \lambda & \lambda^2 \\ 0 & 0 & -\lambda^3-\lambda \end{bmatrix}$$

$$\xrightarrow[\substack{[3,(-\lambda)2] \\ [E(3)(-1)]}]{}\begin{bmatrix} 1 & 0 & 0 \\ 0 & \lambda & 0 \\ 0 & 0 & \lambda^3+\lambda \end{bmatrix}$$

这便是所求的标准形。其中 $d_1(\lambda)=1, d_2(\lambda)=\lambda, d_3(\lambda)=\lambda^3+\lambda$。

□注：记号说明：箭头上方的〔3,(1−λ)2〕表示用(1−λ)乘第二行并加到第三行上；箭头下方的表示相应的列运算，又箭头下方的〔E(3)(−1)〕表示第三列乘以(−1)，余类推。□

我们再回头讨论多项式矩阵的其他一些性质。

定义 3 设多项式矩阵 $A(\lambda)$ 的秩 $r\geqslant 1$，则 $A(\lambda)$ 中的所有非零的 k ($1\leqslant k\leqslant r$) 阶子式的首一最大公因式 $D_k(\lambda)$，称为 $A(\lambda)$ 的 k 阶行列式因子。

定理 3 若 $A(\lambda)\cong B(\lambda)$，则 $A(\lambda)$，$B(\lambda)$ 必有相同的秩及相同的各阶行列式因子。

设 $A(\lambda)$ 经过一次初等变换化为 $B(\lambda)$，则只须就三种初等变换的每一种证明 $A(\lambda)$ 与 $B(\lambda)$ 有相同的秩及相同的各阶行列式因子就行了。但这一证明比较容易，我们把它省略了。

定义 4 在 $A(\lambda)$ 的史密斯标准形 $J(\lambda)$ 中，多项式 $d_1(\lambda), d_2(\lambda), \cdots, d_r(\lambda)$ 称为 $A(\lambda)$ 的不变因式。

由于 $A(\lambda)\cong J(\lambda)$，故由定理 3 可推得：

$D_1(\lambda)=d_1(\lambda), D_2(\lambda)=d_1(\lambda)d_2(\lambda), \cdots,$

$D_r(\lambda)=d_1(\lambda)d_2(\lambda)\cdots d_r(\lambda)$。

从而有：

$$d_1(\lambda)=D_1(\lambda), d_2(\lambda)=\frac{D_2(\lambda)}{D_1(\lambda)}, \cdots, d_r(\lambda)=\frac{D_r(\lambda)}{D_{r-1}(\lambda)}. \qquad (1)$$

这给出了不变因式与行列式因子的关系，通过求行列式因子，也就可以求出 $A(\lambda)$ 的不变因式，从而可得到 $A(\lambda)$ 的史密斯标准形。对于一些比较特殊的矩阵，用这个方法可以较易求得其不变因式。我们在§2已见过不少这方面的例子，这里就不重复了。

由关系式(1)看出，$A(\lambda)$ 的不变因式完全由其各阶行列式因子所唯一确定，所以史密斯标准形是唯一的。

由关系式(1)还看出行列式因子之间满足整除关系：$D_k(\lambda)\mid$

$D_{k+1}(\lambda), k=1,2,\cdots,r-1$。

又当 $A(\lambda)$为 n 阶可逆矩阵时，则因$|A(\lambda)|=c$ 为非零常数，故 $D_n(\lambda)=1$，从而 $D_k(\lambda)=1, d_k(\lambda)=1\ (k=1,2,\cdots,n)$。亦即可逆矩阵 $A(\lambda)$的标准形是单位矩阵 E。反之，与单位矩阵等价的多项式矩阵必可逆(因其行列式等于非零常数)。故 $A(\lambda)$为单模矩阵的充要条件，是 $A(\lambda)$可以表示成初等矩阵的乘积。

本节的多项式矩阵 $A(\lambda)$一直是在任一数域 P 上进行讨论的。如果取复数域 C 替代 P，则我们还可以把 $A(\lambda)$的那些次数大于 1 的不变因式分解为一次因式的方幂的乘积，因而有下述概念：

定义 5 把 $A(\lambda)$的每个次数$\geqslant 1$ 的不变因式 $d_k(\lambda)$分解为互不相同的一次因式的方幂的乘积，所有这些一次因式的方幂(相同的按出现的次数计算)，称为多项式矩阵 $A(\lambda)$的**初级因子**。

对于特殊的多项式矩阵 $\lambda E-A$，我们在§2 已讨论过它的初级因子，目的在于求数字矩阵 A 的约当标准形。在那里，我们是通过计算行列式因子来求不变因式的。现在我们可以应用本节介绍的一般方法来计算 $\lambda E-A$ 的不变因式了。这就是通过初等变换，化多项式矩阵 $\lambda E-A$ 为史密斯标准形的办法。求出不变因式后，再计算出初级因子；便可以写出矩阵 A 的约当标准形了。

前面的例 2 和例 3 所给出的多项式矩阵 $A(\lambda)$，其不变因式分别是

$1,\lambda,\lambda(\lambda-1)(\lambda-2)$；(例 2)

$1,\lambda,\lambda(\lambda^2+1)$。(例 3)

而初级因子则分别是：$\lambda,\lambda,\lambda-1,\lambda-2$ 与 $\lambda,\lambda,\lambda+i,\lambda-i$。

例 4 求矩阵

$$A=\begin{bmatrix}-1 & -2 & 6\\ -1 & 0 & 3\\ -1 & -1 & 4\end{bmatrix}$$

的约当标准形。

解 先求出 $\lambda E-A$ 的不变因式及初级因子，为此应用初等变换化 $\lambda E-A$ 为史密斯标准形。我们有：

$$\lambda E-A=\begin{bmatrix}\lambda+1 & 2 & -6\\ 1 & \lambda & -3\\ 1 & 1 & \lambda-4\end{bmatrix}\to\begin{bmatrix}0 & -\lambda+1 & -\lambda^2+3\lambda-2\\ 0 & \lambda-1 & -\lambda+1\\ 1 & 1 & \lambda-4\end{bmatrix}$$

$$\to\begin{bmatrix}1 & 0 & 0\\ 0 & \lambda-1 & -\lambda+1\\ 0 & -\lambda+1 & -\lambda^2+3\lambda-2\end{bmatrix}\to\begin{bmatrix}1 & 0 & 0\\ 0 & \lambda-1 & -\lambda+1\\ 0 & 0 & -\lambda^2+2\lambda-1\end{bmatrix}$$

$$\to\begin{bmatrix}1 & 0 & 0\\ 0 & \lambda-1 & 0\\ 0 & 0 & (\lambda-1)^2\end{bmatrix}=J(\lambda).$$

由此可见不变因式为 $1,\lambda-1,(\lambda-1)^2$。从而初级因子为 $\lambda-1,(\lambda-1)^2$。故 A 的约当标准形为：

$$J=\begin{bmatrix}1 & 0 & 0\\ 0 & 1 & 0\\ 0 & 1 & 1\end{bmatrix}$$

定理 2 是多项式矩阵的理论中主要定理之一，此外这个理论还有几个定理也是重要的，有的证明相当复杂，在这里，我们只作如下简要的讨论。

定理 4 两个多项式矩阵 $A(\lambda)$ 与 $B(\lambda)$ 等价的充要条件，是两者有相同的行列式因子，或相同的不变因式。

证明 必要性由定理 3 及关系式(1)已经得知；充分性的证明如下：若 $A(\lambda),B(\lambda)$ 有相同的不变因式（或有相同的行列式因子），则 $A(\lambda),B(\lambda)$ 与同一个史密斯标准形等价，从而 $A(\lambda)\cong B(\lambda)$。

数字矩阵相似的条件，也可以由它的特征矩阵来描述。应用多项式矩阵的前述基础知识，可以推出下述结论（证略）：

数域 P 上的两个 n 阶矩阵 A,B 相似的充要条件是它们的特征矩阵 $\lambda E-A$ 与 $\lambda E-B$ 等价。

即是说，$A\sim B\Leftrightarrow\lambda E-A\cong\lambda E-B$。由此易得下面的推论：

$A\sim B\Leftrightarrow A,B$ 有相同的不变因式。如果是在复数域 C 上讨论的，则由上面的推论还可以进一步证明：

$A\sim B\Leftrightarrow A,B$ 有相同的初级因子。

这就是 $A\sim J$（约当标准形）的主要理论依据。

§5 多项式矩阵的互质性与既约性

在系统分析中，常会涉及到多项式矩阵互质性的判别问题。关于可控性问题的讨论，常归结为具有相同行数的多项式矩阵的左互质问题，而可观测性问题的讨论就归结为具有相同列数的多项式矩阵的右互质问题。另外在许多问题中还需以满足既约性为条件。所以，建立一些互质性与既约性的判别准则是很必要的。

从本节起，我们所讲到的多项式矩阵、分式矩阵都是在复数域中进行讨论的。

定义 1 多项式方阵 $R(\lambda)$称为具有相同列数的两个多项式矩阵 $N(\lambda)$与 $D(\lambda)$的一个右公因子，如果存在多项式矩阵$\overline{N}(\lambda)$和$\overline{D}(\lambda)$，使得

$$N(\lambda)=\overline{N}(\lambda)R(\lambda),\quad D(\lambda)=\overline{D}(\lambda)R(\lambda).$$

类似地可以定义左公因子。

定义 2 多项式方阵 $R(\lambda)$称为具有相同列数的两个多项式矩阵 $N(\lambda)$与 $D(\lambda)$的一个最大右公因子（记为 gcrd），如果：

1) $R(\lambda)$是 $N(\lambda)$与 $D(\lambda)$的右公因子；

2) $N(\lambda)$与 $D(\lambda)$的任一其他的右公因子 $R_1(\lambda)$都是 $R(\lambda)$的右乘因子，即有多项式矩阵 $W(\lambda)$使得 $R(\lambda)=W(\lambda)R_1(\lambda)$。

对任意的 $n\times n$ 与 $m\times n$ 的多项式矩阵 $D(\lambda)$与 $N(\lambda)$，它们的 gcrd 都存在。因为

$$R(\lambda)=(D^T(\lambda),N^T(\lambda))^T$$

便是一个。

以下提到的 gcrd 均指 $n\times n$ 方阵。利用多项式矩阵的史密斯标准形，可以证明 $n\times n$ 的 gcrd 的存在性。

定理 1(gcrd 的构造定理)　如果可以找到一个 $(n+m)\times(n+m)$ 的单模矩阵 $G(\lambda)$，使得

$$G(\lambda)\begin{pmatrix}D(\lambda)\\N(\lambda)\end{pmatrix}=\begin{pmatrix}G_{11}(\lambda) & G_{12}(\lambda)\\G_{21}(\lambda) & G_{22}(\lambda)\end{pmatrix}\begin{pmatrix}D(\lambda)\\N(\lambda)\end{pmatrix}=\begin{pmatrix}R(\lambda)\\0\end{pmatrix} \qquad (1)$$

则 $n\times n$ 多项式矩阵 $R(\lambda)$ 即为 $D(\lambda)$ 与 $N(\lambda)$ 的一个 gcrd。

证明　先证 $R(\lambda)$ 是右公因子。为此，把 $G(\lambda)$ 的逆矩阵 $G^{-1}(\lambda)=F(\lambda)$ 写成分块矩阵：

$$G^{-1}(\lambda)=F(\lambda)=\begin{pmatrix}F_{11}(\lambda) & F_{12}(\lambda)\\F_{21}(\lambda) & F_{22}(\lambda)\end{pmatrix}$$

以 $G^{-1}(\lambda)$ 左乘(1)式即得：

$$\begin{pmatrix}D(\lambda)\\N(\lambda)\end{pmatrix}=\begin{pmatrix}F_{11}(\lambda) & F_{12}(\lambda)\\F_{21}(\lambda) & F_{22}(\lambda)\end{pmatrix}\begin{pmatrix}R(\lambda)\\0\end{pmatrix}$$

$$=\begin{pmatrix}F_{11}(\lambda)R(\lambda)\\F_{21}(\lambda)R(\lambda)\end{pmatrix}$$

这表明存在多项式矩阵 $F_{11}(\lambda)$ 及 $F_{21}(\lambda)$，使得

$$D(\lambda)=F_{11}(\lambda)R(\lambda),\quad N(\lambda)=F_{21}(\lambda)R(\lambda)。$$

从而 $R(\lambda)$ 是 $D(\lambda)$ 与 $N(\lambda)$ 的右公因子。

现设 $R_1(\lambda)$ 为另一右公因子，下证 $R_1(\lambda)$ 是 $R(\lambda)$ 的右乘因子。因 $R_1(\lambda)$ 为 $D(\lambda)$，$N(\lambda)$ 的右公因子，故有：

$$D(\lambda)=D_1(\lambda)R_1(\lambda),\quad N(\lambda)=N_1(\lambda)R_1(\lambda) \qquad (2)$$

而由(1)式又有：

$$R(\lambda)=G_{11}(\lambda)D(\lambda)+G_{12}(\lambda)N(\lambda)\ 。 \qquad (3)$$

把(2)式代入(3)式得

$$R(\lambda)=[G_{11}(\lambda)D_1(\lambda)+G_{12}(\lambda)N_1(\lambda)]R_1(\lambda)$$
$$=W(\lambda)R_1(\lambda)。$$

即 $R_1(\lambda)$是 $R(\lambda)$的右乘因子。证毕。

由于单模矩阵都可以表示成一些初等矩阵的乘积，故对一个多项式矩阵左乘一个单模矩阵，相当于对它施行一系列的初等行变换运算。故由上述定理可知，多项式矩阵 $D(\lambda)$与 $N(\lambda)$的一个 gcrd$R(\lambda)$，可通过对矩阵

$$M(\lambda)=\begin{pmatrix} D(\lambda) \\ N(\lambda) \end{pmatrix}$$

施行一些初等行变换来得到，而相应的初等矩阵的乘积就是所要找的单模矩阵 $G(\lambda)$。

例1 设

$$D(\lambda)=\begin{pmatrix} \lambda & 3\lambda+1 \\ -1 & \lambda^2+\lambda-2 \end{pmatrix},\quad N(\lambda)=(-1,\lambda^2+2\lambda-1),$$

求 gcrd$R(\lambda)$。

解

$$\begin{pmatrix} D(\lambda) \\ N(\lambda) \end{pmatrix}=\begin{bmatrix} \lambda & 3\lambda+1 \\ -1 & \lambda^2+\lambda-2 \\ -1 & \lambda^2+2\lambda-1 \end{bmatrix}$$

$$\xrightarrow{P_1:[1,2]}\begin{bmatrix} -1 & \lambda^2+\lambda-2 \\ \lambda & 3\lambda+1 \\ -1 & \lambda^2+2\lambda-1 \end{bmatrix}\xrightarrow[P_4:[(-1)\times(1)]]{P_2:[2,\ \lambda(1)]\quad P_3:[3,(-1)(1)]}\begin{bmatrix} 1 & -\lambda^2-\lambda+2 \\ 0 & \lambda^3+\lambda^2+\lambda+1 \\ 0 & \lambda+1 \end{bmatrix}$$

$$\xrightarrow{P_5:[2,3]}\begin{bmatrix} 1 & -\lambda^2-\lambda+2 \\ 0 & \lambda+1 \\ 0 & \lambda^3+\lambda^2+\lambda-1 \end{bmatrix}\xrightarrow{P_6:[3,-(\lambda^2+1)(2)]}\begin{bmatrix} 1 & -\lambda^2-\lambda+2 \\ 0 & \lambda+1 \\ 0 & 0 \end{bmatrix}$$

因此求得

$$R(\lambda)=\begin{pmatrix} 1 & -\lambda^2-\lambda+2 \\ 0 & \lambda+1 \end{pmatrix}$$

而相应的单模矩阵为

$$G(\lambda)=P_6P_5P_4P_3P_2P_1$$

$$=\begin{bmatrix}1 & 0 & 0\\0 & 1 & 0\\0 & -(\lambda^2+1) & 1\end{bmatrix}\cdot\begin{bmatrix}1 & 0 & 0\\0 & 0 & 1\\0 & 1 & 0\end{bmatrix}\cdot\begin{bmatrix}-1 & 0 & 0\\0 & 1 & 0\\0 & 0 & 1\end{bmatrix}$$

$$\begin{bmatrix}1 & 0 & 0\\0 & 1 & 0\\-1 & 0 & 1\end{bmatrix}\cdot\begin{bmatrix}1 & 0 & 0\\\lambda & 1 & 0\\0 & 0 & 1\end{bmatrix}\cdot\begin{bmatrix}0 & 1 & 0\\1 & 0 & 0\\0 & 0 & 1\end{bmatrix}$$

$$=\begin{bmatrix}0 & -1 & 0\\0 & -1 & 1\\1 & \lambda^2+\lambda+1 & -(\lambda^2+1)\end{bmatrix}$$

gcrd 的基本性质：

1) 不唯一性。若 $R(\lambda)$为具有相同列数 p 的两个多项式矩阵 $D(\lambda)$与 $N(\lambda)$的一个 gcrd，而 $W(\lambda)$为任一 p 阶单模矩阵，则 $W(\lambda)R(\lambda)$也是 $D(\lambda)$和 $N(\lambda)$的一个 gcrd。

2) 若 $R_1(\lambda)$与 $R_2(\lambda)$是 $D(\lambda)$与 $N(\lambda)$的任意两个 gcrd，则当 $R_1(\lambda)$为满秩矩阵或单模矩阵时，$R_2(\lambda)$也一定是满秩矩阵或单模矩阵。

3) 对给定的 $n\times n$ 与 $m\times n$ 多项式矩阵 $D(\lambda)$与 $N(\lambda)$，则当

$$\begin{pmatrix}D(\lambda)\\N(\lambda)\end{pmatrix}\text{为列满秩即 rank}\begin{pmatrix}D(\lambda)\\N(\lambda)\end{pmatrix}=n$$

时，$D(\lambda)$与 $N(\lambda)$的所有 gcrd 都必定是满秩的。

4) 若 $R(\lambda)$是 $n\times n$ 与 $m\times n$ 多项式矩阵 $D(\lambda)$与 $N(\lambda)$的一个 gcrd，则 $R(\lambda)$可表示为

$$R(\lambda)=X(\lambda)D(\lambda)+Y(\lambda)N(\lambda),$$

其中 $X(\lambda)$，$Y(\lambda)$分别是 $n\times n$ 与 $n\times m$ 多项式矩阵。

证明 由构造定理，即得

$$R(\lambda)=G_{11}(\lambda)D(\lambda)+G_{12}(\lambda)N(\lambda),$$

取 $X(\lambda)=G_{11}(\lambda)$，$Y(\lambda)=G_{12}(\lambda)$便得所证。

其他三个性质的证明从略。又 1)与 3)也是借助构造定理来证

明的，可见该定理的重要性。

同定义 2 相类似，也可以引入两个多项式矩阵的最大左公因子(gcld)的概念，建立其相应的构造定理及基本性质。从略。

现转移到多项式矩阵的互质性问题。

定义 3 两个具有相同列数的多项式矩阵 $D(\lambda)$ 与 $N(\lambda)$ 称为是右互质的，如果它们的最大右公因子(gcrd)为单模矩阵。

类似地可定义左互质概念。注意右(左)互质时，未必是左(右)互质的。

定理 2(贝佐特判别准则) 两个 $n\times n$ 与 $m\times n$ 多项式矩阵 $D(\lambda)$ 与 $N(\lambda)$ 为右互质的充要条件，是存在两个 $n\times n$ 与 $n\times m$ 多项式矩阵 $X(\lambda)$ 与 $Y(\lambda)$，使得下面的贝佐特(Bezout)等式成立：

$$X(\lambda)D(\lambda)+Y(\lambda)N(\lambda)=E. \qquad (1)$$

证明 必要性。设 $D(\lambda)$ 与 $N(\lambda)$ 是右互质的，因而它们的 gcrd$R(\lambda)$ 为单模矩阵。由构造定理即得

$$R(\lambda)=G_{11}(\lambda)D(\lambda)+G_{12}(\lambda)N(\lambda). \qquad (2)$$

因 $R^{-1}(\lambda)$ 存在且为多项式矩阵，以 $R^{-1}(\lambda)$ 左乘(2)式即得

$$R^{-1}(\lambda)G_{11}(\lambda)D(\lambda)+R^{-1}(\lambda)G_{12}(\lambda)N(\lambda)=E. \qquad (3)$$

令

$$X(\lambda)=R^{-1}(\lambda)G_{11}(\lambda),Y(\lambda)=R^{-1}(\lambda)G_{12}(\lambda),$$

则(3)式便为(1)式。

充分性。设(1)式成立，令 $R(\lambda)$ 为 $D(\lambda)$ 与 $N(\lambda)$ 的一个 gcrd，则存在多项式矩阵 $\overline{D}(\lambda)$ 及 $\overline{N}(\lambda)$，使得

$$D(\lambda)=\overline{D}(\lambda)R(\lambda), \quad N(\lambda)=\overline{N}(\lambda)R(\lambda). \qquad (4)$$

将(4)代入(1)可得

$$[X(\lambda)\overline{D}(\lambda)+Y(\lambda)\overline{N}(\lambda)]R(\lambda)=E. \qquad (5)$$

因方括号内部分是个多项式矩阵，故由单模矩阵定义及(5)式，便知 $R(\lambda)$ 为单模矩阵。从而 $D(\lambda)$ 与 $N(\lambda)$ 是右互质的。证毕。

利用多项式矩阵的史密斯标准形，也可以判别两个多项式矩

阵的互质性。即是说我们有下述定理。

定理 3 两个 $n\times n$ 与 $m\times n$ 多项式矩阵 $D(\lambda)$ 与 $N(\lambda)$ 为右互质的充要条件，是矩阵

$$\begin{pmatrix} D(\lambda) \\ N(\lambda) \end{pmatrix}$$

的史密斯标准形为

$$\begin{bmatrix} E \\ 0 \end{bmatrix}$$

而两个 $m\times m$ 与 $m\times n$ 多项式矩阵 $A(\lambda)$ 与 $B(\lambda)$ 为左互质的充要条件，是矩阵 $(A(\lambda),B(\lambda))$ 的史密斯标准形为 $(E,0)$。

证明 这里只证明右互质的结论. 由构造定理即有

$$G(\lambda)\begin{pmatrix} D(\lambda) \\ N(\lambda) \end{pmatrix} = \begin{pmatrix} R(\lambda) \\ 0 \end{pmatrix} = \begin{bmatrix} E \\ 0 \end{bmatrix} R(\lambda), \qquad (1)$$

其中 $G(\lambda)$ 是单模矩阵。

若 $D(\lambda),N(\lambda)$ 为右互质的，则由定义 3，便知道它们的最大右公因子 $R(\lambda)$ 是单模矩阵，而 $R^{-1}(\lambda)$ 也是单模矩阵. 将(1)式右乘 $R^{-1}(\lambda)$ 即得

$$G(\lambda)\begin{pmatrix} D(\lambda) \\ N(\lambda) \end{pmatrix} R^{-1}(\lambda) = \begin{bmatrix} E \\ 0 \end{bmatrix} R(\lambda)R^{-1}(\lambda) = \begin{bmatrix} E \\ 0 \end{bmatrix} \qquad (2)$$

如记单模矩阵 $R^{-1}(\lambda)$ 为 $R_1(\lambda)$，则(2)式即为

$$G(\lambda)\begin{pmatrix} D(\lambda) \\ N(\lambda) \end{pmatrix} R_1(\lambda) = \begin{bmatrix} E \\ 0 \end{bmatrix} \qquad (3)$$

因此

$$\begin{pmatrix} D(\lambda) \\ N(\lambda) \end{pmatrix} \cong \begin{bmatrix} E \\ 0 \end{bmatrix}, \text{记为 } M(\lambda) \cong J \ . \qquad (4)$$

从而有相同的不变因式(而 J 的不变因式全为 1)，故有相同的史密斯标准形，但 J 的标准形就是自身，故 $M(\lambda)$ 的标准形是 J。故必

要性得证。

再证充分性。设 $M(\lambda)$ 的标准形是 J(见(4)式)。这说明存在单模矩阵 $G(\lambda)$ 及 $R_1(\lambda)$，使得(3)式成立。但 $R_1^{-1}(\lambda)$ 也是单模矩阵，以 $R_1^{-1}(\lambda)$ 右乘(3)式，即得

$$G(\lambda)\begin{bmatrix}D(\lambda)\\N(\lambda)\end{bmatrix}=\begin{bmatrix}R_1^{-1}(\lambda)\\0\end{bmatrix}$$

由定理 1，这等式表明单模矩阵 $R_1^{-1}(\lambda)$ 是 $D(\lambda)$ 与 $N(\lambda)$ 的一个 gcrd，再由定义 3 即知 $D(\lambda)$ 与 $N(\lambda)$ 是右互质的。证毕。

最后讨论多项式矩阵的既约性问题。

设 $M(\lambda)=(m_{ij}(\lambda))_{p\times q}$ 是个多项式矩阵。定义：

$$K_{ri}=\max_{1\leqslant j\leqslant q}\{\deg(m_{ij}(\lambda))\},$$

$$K_{cj}=\max_{1\leqslant i\leqslant p}\{\deg(m_{ij}(\lambda))\}.$$

前者称为 $M(\lambda)$ 的第 i 个行次数，后者叫做 $M(\lambda)$ 的第 j 个列次数。并分别记为 $\delta_{ri}M(\lambda)$ 及 $\delta_{cj}M(\lambda)$。

例 2 若

$$M(\lambda)=\begin{bmatrix}\lambda^2+4\lambda-2 & 2\lambda^2+5\lambda & \lambda+6\\ \lambda+3 & \lambda^3-2\lambda & 4\end{bmatrix}$$

则有：

$$K_{r1}=2,\quad K_{r2}=3;$$

$$K_{c1}=2,\quad K_{c2}=3,\quad K_{c3}=1.$$

借助行次数及列次数的概念，多项式矩阵 $M(\lambda)$ 可以表示为下述两种形式。

列次表示式：

$$M(\lambda)=M_{hc}H_c(\lambda)+M_{lc}(\lambda), \tag{1}$$

这里 $H_c(\lambda)=\mathrm{diag}\{\lambda^{K_{c1}},\lambda^{K_{c2}},\cdots,\lambda^{K_{cq}}\}$；$M_{hc}$ 为 $p\times q$ 的数字矩阵，它的第 j 列为 $M(\lambda)$ 的第 j 列中相应于 $\lambda^{K_{cj}}$ 的系数组成的列；M_{hc} 称为列次系数矩阵。又 $M_{lc}(\lambda)$ 为低次剩余多项式矩阵，且

$$\delta_{cj} M_{lc}(\lambda) < K_{cj} \quad (j = 1, 2, \cdots, q).$$

例 3 设

$$M(\lambda) = \begin{bmatrix} \lambda^2 + 4\lambda + 1 & 2\lambda^2 & \lambda + 4 \\ \lambda + 3 & 4\lambda^3 - 2\lambda & 3\lambda + 1 \end{bmatrix}$$

则其列次表示式为

$$M(\lambda) = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 4 & 3 \end{bmatrix} \begin{bmatrix} \lambda^2 & & \\ & \lambda^3 & \\ & & \lambda \end{bmatrix} + \begin{bmatrix} 4\lambda + 1 & 2\lambda^2 & 4 \\ \lambda + 3 & -2\lambda & 1 \end{bmatrix}$$

如果限于考虑 $M(\lambda)$ 为方阵即 $q=p$ 的情形，则由(1)式还可导出一个有用的关系式($M(\lambda)$ 的行列式表示)：

$$|M(\lambda)| = |M_{hc}| \cdot \lambda^{\Sigma K_{cj}} + \text{次数低于 } \Sigma K_{cj} \text{ 的各项}.$$

类似地有 $M(\lambda)$ 的行次表示式：

$$M(\lambda) = H_r(\lambda) M_{hr} + M_{lr}(\lambda), \qquad (2)$$

这里 $H_r(\lambda) = \text{diag}\{\lambda^{K_{r1}}, \lambda^{K_{r2}}, \cdots, \lambda^{K_{rp}}\}$；$M_{hr}$ 为 $p \times q$ 数字矩阵，称为行次系数矩阵，它的第 i 行即为 $M(\lambda)$ 的第 i 行中相应于 $\lambda^{K_{ri}}$ 的系数组成的行；而 $M_{lr}(\lambda)$ 为低次剩余多项式矩阵，且满足 $\delta_{ri} M_{lr}(\lambda) < K_{ri}$ $(i = 1, 2, \cdots, p)$。

又当 $M(\lambda)$ 为 $p \times p$ 方阵时，由(2)式还可导出相应的行列式表示式：

$$|M(\lambda)| = |M_{hr}| \lambda^{\Sigma K_{ri}} + \text{次数低于 } \Sigma K_{ri} \text{ 的各项}.$$

定义 4 设 $M(\lambda)$ 为满秩 p 阶多项式矩阵，如果

$$\deg |M(\lambda)| = \sum_{j=1}^{p} \delta_{cj} M(\lambda) = K_{c1} + K_{c2} + \cdots + K_{cp}$$

则称 $M(\lambda)$ 是列既约的；如果

$$\deg |M(\lambda)| = \sum_{i=1}^{p} \delta_{ri} M(\lambda) = K_{r1} + K_{r2} + \cdots + K_{rp}$$

则称 $M(\lambda)$ 是行既约的。

例 4 设有满秩多项式矩阵

$$M(\lambda)=\begin{pmatrix}3\lambda^2+2\lambda & 2\lambda+4\\ \lambda^2+\lambda-3 & 7\lambda\end{pmatrix}$$

则有：

$$\deg|M(\lambda)|=3,$$

$$K_{c1}=2,\quad K_{c2}=1,\quad \Sigma K_{cj}=3,$$

$$K_{r1}=2,\quad K_{r2}=2,\quad \Sigma K_{ri}=4。$$

因此，这个多项式矩阵是列既约的，但不是行既约的。

定理 4 若 $M(\lambda)$ 为 p 阶满秩多项式矩阵，则：

1) $M(\lambda)$ 为列既约 $\Leftrightarrow M_{hc}$ 为满秩的；

2) $M(\lambda)$ 为行既约 $\Leftrightarrow M_{hr}$ 为满秩的。

证明 由于

$$|M(\lambda)|=|M_{hc}|\cdot\lambda^{\Sigma K_{cj}}+\text{低次项},$$

故当且只当 $|M_{hc}|\neq0$ 即 M_{hc} 为满秩时，始有

$$\deg|M(\lambda)|=\Sigma K_{cj},$$

亦即 $M(\lambda)$ 为列既约的。因而 1)得证。同理可证 2)。证毕。

例 5 对于例 4 的矩阵 $M(\lambda)$，可求得其列次表示及行次表示为：

列次表示 $M(\lambda)=\begin{pmatrix}3 & 2\\ 1 & 7\end{pmatrix}\begin{pmatrix}\lambda^2 & 0\\ 0 & \lambda\end{pmatrix}+\begin{pmatrix}2\lambda & 4\\ \lambda-3 & 0\end{pmatrix}$

行次表示 $M(\lambda)=\begin{pmatrix}\lambda^2 & 0\\ 0 & \lambda^2\end{pmatrix}\begin{pmatrix}3 & 0\\ 1 & 0\end{pmatrix}+\begin{pmatrix}2\lambda & 2\lambda+4\\ \lambda-3 & 7\lambda\end{pmatrix}$

因此有

$$M_{hc}=\begin{pmatrix}3 & 2\\ 1 & 7\end{pmatrix},\qquad M_{hr}=\begin{pmatrix}3 & 0\\ 1 & 0\end{pmatrix}$$

故由 M_{hc} 满秩知 $M(\lambda)$ 是列既约的，但因 M_{hr} 不是满秩的，故 $M(\lambda)$ 不是行既约的。

如何把一个非既约多项式矩阵化为既约的，这在系统分析中

是时常出现的。应用初等变换的办法便可达到这个目的。设 $M(\lambda)$是给定的 p 阶满秩非既约多项式矩阵，则可以找到 p 阶单模矩阵 $G(\lambda)$及 $F(\lambda)$，使得 $M(\lambda)G(\lambda)$与 $F(\lambda)M(\lambda)$为列既约（或行既约）的。这种做法的实质，就是通过对 $M(\lambda)$进行适当的列（或行）初等变换，来降低它的某些列（行）次数，以满足既约性定义 4 中所表达的要求。

例6 设

$$M(\lambda)=\begin{bmatrix}(\lambda+2)^2(\lambda+3)^2 & -(\lambda+2)^2(\lambda+3)\\ 0 & \lambda+3\end{bmatrix}$$

则 $M(\lambda)$满秩但非列既约。

我们有单模矩阵

$$G(\lambda)=\begin{pmatrix}1 & 0\\ \lambda+3 & 1\end{pmatrix},\qquad F(\lambda)=\begin{pmatrix}1 & (\lambda+2)^2\\ 0 & 1\end{pmatrix}$$

而使得

$$M(\lambda)G(\lambda)=\begin{pmatrix}0 & -(\lambda+2)^2(\lambda+3)\\ (\lambda+3)^2 & \lambda+3\end{pmatrix}$$

及

$$F(\lambda)M(\lambda)=\begin{pmatrix}(\lambda+2)^2(\lambda+3)^2 & 0\\ 0 & \lambda+3\end{pmatrix}$$

都是列既约的。

§6 有理分式矩阵的标准形及其仿分式分解

由于应用上的需要，把多项式矩阵推广为有理分式矩阵是必要的。如果矩阵 $G(\lambda)=(g_{ij}(\lambda))_{m\times n}$的元素

$$g_{ij}(\lambda)=\frac{a_{ij}(\lambda)}{b_{ij}(\lambda)}\quad(i=1,2,\cdots,m;j=1,2,\cdots,n)$$

都是 λ 的有理分式（这里 $a_{ij}(\lambda)$与 $b_{ij}(\lambda)$都是 λ 的多项式），则 $G(\lambda)$称为**有理分式矩阵**，简称**分式矩阵**。显然，多项式矩阵是它的特例。

由于有理分式经过四则运算后仍为有理分式，因而可以像数字矩阵那样类似地定义 $G(\lambda)$的各种运算及概念，如 $G(\lambda)$的子式、秩，等等。当 $G(\lambda)$为方阵时，如其行列式$|G(\lambda)|\neq 0$，则称 $G(\lambda)$是可逆的，其逆阵记为 $G^{-1}(\lambda)$，它也是个有理分式矩阵，当然仍要求满足条件 $G(\lambda)G^{-1}(\lambda)=E$。

有理分式矩阵在线性系统理论中是个重要工具．传递函数矩阵都是有理分式矩阵，它的理论其实就是分式矩阵的理论。本节只介绍分式矩阵的一种标准形及其仿分式分解．

定理 1 设 $G(\lambda)=(g_{ij}(\lambda))_{m\times n}\neq 0$（矩阵）是有理分式矩阵，且 $\text{rank}G(\lambda)=r(\geqslant 1)$，则存在 $m\times n$ 单模多项式矩阵 $P(\lambda)$及 $n\times n$ 单模多项式矩阵 $Q(\lambda)$，使得

$$P(\lambda)G(\lambda)Q(\lambda) = M(\lambda) = \begin{pmatrix} T(\lambda) & \\ & 0 \end{pmatrix}$$

其中

$$T(\lambda) = \text{diag}\left(\frac{\varphi_1(\lambda)}{t_1(\lambda)}, \frac{\varphi_2(\lambda)}{t_2(\lambda)}, \cdots, \frac{\varphi_r(\lambda)}{t_r(\lambda)}\right)$$

$\varphi_i(\lambda)$，$t_i(\lambda)$都是首一多项式，且满足条件：

1） $t_i(\lambda)$与 $\varphi_i(\lambda)$互质（$i=1,2,\cdots,r$）；

2） $\varphi_i(\lambda)|\varphi_{i+1}(\lambda)\quad(i=1,2,\cdots,r-1)$；

3） $t_{i+1}(\lambda)|t_i(\lambda)\quad(i=1,2,\cdots,r-1)$。

又 $M(\lambda)$称为有理分式矩阵 $G(\lambda)$的史密斯—麦克米伦（Smith—Mcmillan）标准形。它是 B. Mcmillan 于 1952 年提出来的。它的意义在于为系统分析，特别是分析多变量系统的极点和零点提供了一种重要的概念性和理论性工具。

证明 设有理分式矩阵

$$G(\lambda) = (g_{ij}(\lambda))_{m\times n} = \left(\frac{a_{ij}(\lambda)}{b_{ij}(\lambda)}\right)_{m\times n}$$

的元素中 $m\times n$ 个分母多项式

$$b_{ij}(\lambda) \quad (i=1,2,\cdots,m;\quad j=1,2,\cdots,n)$$

的最小公倍为 $b(\lambda)$（且为首一多项式），则 $b(\lambda)G(\lambda)$ 是多项式矩阵，且与 $G(\lambda)$ 有相同的秩 r。故有 $m\times n$ 单模矩阵 $P(\lambda)$ 及 $n\times n$ 单模矩阵 $Q(\lambda)$，把 $b(\lambda)G(\lambda)$ 化为史密斯标准形，即

$$P(\lambda)(b(\lambda)G(\lambda))Q(\lambda)=\begin{bmatrix}T_1(\lambda) & \\ & 0\end{bmatrix}$$

这里 $T_1(\lambda)=\mathrm{diag}(d_1(\lambda),d_2(\lambda),\cdots,d_r(\lambda))$，而且 $d_i(\lambda)\mid d_{i+1}(\lambda)$ $(i=1,2,\cdots,r-1)$。因此

$$P(\lambda)G(\lambda)Q(\lambda)=\begin{bmatrix}\dfrac{T_1(\lambda)}{b(\lambda)} & \\ & 0\end{bmatrix} \tag{1}$$

而

$$\frac{T_1(\lambda)}{b(\lambda)}=\mathrm{diag}\left(\frac{d_1(\lambda)}{b(\lambda)},\frac{d_2(\lambda)}{b(\lambda)},\cdots,\frac{d_r(\lambda)}{b(\lambda)}\right)$$

把每个 $d_i(\lambda)/b(\lambda)$ 化为既约分式 $\varphi_i(\lambda)/l_i(\lambda)$（且 $\varphi_i(\lambda)$ 与 $l_i(\lambda)$ 均为首一多项式），为简单起见，(1)式经这样处理后左边仍用原式，于是得

$$P(\lambda)G(\lambda)Q(\lambda)=\begin{bmatrix}T(\lambda) & \\ & 0\end{bmatrix}$$

其中

$$T(\lambda)=\mathrm{diag}\left(\frac{\varphi_1(\lambda)}{l_1(\lambda)},\frac{\varphi_2(\lambda)}{l_2(\lambda)},\cdots,\frac{\varphi_r(\lambda)}{l_r(\lambda)}\right)$$

又因 $\varphi_i(\lambda)/l_i(\lambda)$ 为既约分式，所以 $\varphi_i(\lambda)$ 与 $l_i(\lambda)$ 互质 $(i=1,2,\cdots,r)$，即定理中的 1)得证。而 2)与 3)也不难证明，从略。定理证毕。

上述标准形是唯一的。

例 1 求有理分式矩阵

$$G(\lambda)=\begin{bmatrix}\frac{1-\lambda}{\lambda(\lambda+1)} & \frac{\lambda}{\lambda+1} & \frac{1}{\lambda+1}\\ \frac{1}{\lambda+1} & \frac{1}{\lambda+1} & -\frac{1}{\lambda+1}\\ \frac{\lambda^2+1}{\lambda(\lambda+1)} & \frac{\lambda}{\lambda+1} & -\frac{\lambda}{\lambda+1}\end{bmatrix}$$

的标准形。

解 $G(\lambda)$的元素中分母多项式的最小公倍式为$b(\lambda)=\lambda(\lambda+1)$，而

$$b(\lambda)G(\lambda)=\begin{bmatrix}1-\lambda & \lambda^2 & \lambda\\ \lambda & \lambda & -\lambda\\ \lambda^2+1 & \lambda^2 & -\lambda^2\end{bmatrix}$$

$$\cong\begin{bmatrix}1 & 0 & 0\\ 0 & \lambda & 0\\ 0 & 0 & \lambda(\lambda+1)\end{bmatrix}$$

所以

$$G(\lambda)\cong\begin{bmatrix}\frac{1}{\lambda(\lambda+1)} & 0 & 0\\ 0 & \frac{1}{\lambda+1} & 0\\ 0 & 0 & 1\end{bmatrix}$$

定义 1 设$G(\lambda)$是$m\times n$有理分式矩阵，如果存在$m\times m$单模（多项式）矩阵$P_1(\lambda)$及$m\times n$多项式矩阵$Q_1(\lambda)$，使得

$$G(\lambda)=P_1^{-1}(\lambda)Q_1(\lambda),\tag{1}$$

则(1)式称为$G(\lambda)$的一个左分解，或$G(\lambda)$的一个左矩阵分式描述；而当$P_1(\lambda)$与$Q_1(\lambda)$为左互质时，则(1)式称为$G(\lambda)$的一个左既约分解。

若存在$n\times n$单模（多项式）矩阵$P_2(\lambda)$及$m\times n$多项式矩阵$Q_2(\lambda)$，使得

$$G(\lambda)=Q_2(\lambda)P_2^{-1}(\lambda),\tag{2}$$

则(2)式称为$G(\lambda)$的一个右分解，或$G(\lambda)$的一个右矩阵分式描述；若$P_2(\lambda)$与$Q_2(\lambda)$为右互质的，则(2)式称为$G(\lambda)$的一个右既约分解。

定理 2 任何$m\times n$有理分式矩阵$G(\lambda)$都存在左分解、右分解、左既约分解及右既约分解。

证明 由定理1即有

$$P(\lambda)G(\lambda)Q(\lambda)=\begin{pmatrix}T(\lambda) & \\ & 0\end{pmatrix},\quad (T(\lambda)\text{ 如上述})$$

因此

$$\begin{aligned}G(\lambda)&=P^{-1}(\lambda)\begin{pmatrix}T(\lambda) & \\ & 0\end{pmatrix}Q^{-1}(\lambda)\\&=P^{-1}(\lambda)\begin{pmatrix}T_0(\lambda) & \\ & E_{m-r}\end{pmatrix}\begin{pmatrix}\Phi(\lambda) & \\ & 0\end{pmatrix}Q^{-1},\end{aligned}$$

其中

$$T_0(\lambda)=\mathrm{diag}(t_1^{-1}(\lambda),t_2^{-1}(\lambda),\cdots,t_r^{-1}(\lambda)),$$
$$\Phi(\lambda)=\mathrm{diag}(\varphi_1(\lambda),\varphi_2(\lambda),\cdots,\varphi_r(\lambda)).$$

取

$$P_1(\lambda)=\begin{pmatrix}T_0^{-1}(\lambda) & \\ & E_{m-r}\end{pmatrix}P(\lambda),$$

$$Q_1(\lambda)=\begin{bmatrix}\Phi(\lambda) & \\ & 0\end{bmatrix}Q^{-1}(\lambda).$$

则有

$$G(\lambda)=P_1^{-1}(\lambda)Q_1(\lambda),$$

因而得到$G(\lambda)$的一个左分解。

同样有

$$G(\lambda)=P^{-1}(\lambda)\begin{bmatrix}T(\lambda) & \\ & 0\end{bmatrix}Q^{-1}(\lambda)$$

$$=P^{-1}(\lambda)\begin{bmatrix}\Phi(\lambda) & \\ & 0\end{bmatrix}\begin{pmatrix}T_0(\lambda) & \\ & E_{n-r}\end{pmatrix}Q^{-1}(\lambda)$$

取

$$Q_2(\lambda)=P^{-1}(\lambda)\begin{bmatrix}\Phi(\lambda) & \\ & 0\end{bmatrix}$$

$$P_2(\lambda)=Q(\lambda)\begin{pmatrix}T_0^{-1}(\lambda) & \\ & E_{n-r}\end{pmatrix}$$

则有

$$G(\lambda)=Q_2(\lambda)P_2^{-1}(\lambda).$$

这便是 $G(\lambda)$的一个右分解.

若 $G(\lambda)=P_1^{-1}(\lambda)Q_1(\lambda)$是左分解,则可令 $R(\lambda)$是 $P_1(\lambda)$与 $Q_1(\lambda)$的最大左公因子,于是由

$$P_1(\lambda)=R(\lambda)P(\lambda),\quad Q_1(\lambda)=R(\lambda)Q(\lambda)$$

可知 $P(\lambda)$,$Q(\lambda)$左互质,但由于 $R(\lambda)$可逆,故

$$P_1^{-1}(\lambda)Q_1(\lambda)=P^{-1}(\lambda)Q(\lambda).$$

所以 $P^{-1}(\lambda)Q(\lambda)$是 $G(\lambda)$的左既约分解.

仿此可证 $G(\lambda)$存在右既约分解. 证毕.

有理分式矩阵的左、右分解是不唯一的.

例 2 求传递函数矩阵

$$G(s)=\begin{bmatrix}\dfrac{s+1}{(s+2)^2(s+3)^2} & \dfrac{s}{(s+3)^2}\\[2mm] \dfrac{-(s+1)}{(s+3)^2} & \dfrac{-s}{(s+3)^2}\end{bmatrix}$$

的一个右分解及一个左分解.

解 容易求出 $G(s)$的列的最小公分母依次为

$$d_{c1}(s)=(s+2)^2(s+3)^2,\quad d_{c2}(s)=(s+3)^2。$$

因此得 $G(s)$的一个右分解:

$$G(s)=\begin{pmatrix}s+1 & s\\ -(s+1)^2(s+2)^2 & -s\end{pmatrix}\begin{pmatrix}(s+2)^2(s+3)^2 & \\ & (s+3)^2\end{pmatrix}^{-1}$$

又 $G(s)$的行最小公分母为：

$$d_{r1}(s)=(s+2)^2(s+3)^2,\quad d_{r2}(s)=(s+3)^2。$$

所以 $G(s)$的一个左分解为：

$$G(s)=\begin{pmatrix}(s+2)^2(s+3)^2 & \\ & (s+3)^2\end{pmatrix}^{-1}\begin{pmatrix}s+1 & s(s+2)^2\\ -(s+1) & -s\end{pmatrix}$$

□注：例 2 的解法中用到下述运算规则：

(1)
$$\begin{bmatrix}a_{11} & a_{12} & \cdots & a_{1n}\\ a_{21} & a_{22} & \cdots & a_{2n}\\ \cdots & \cdots & \cdots & \cdots\\ a_{m1} & a_{m2} & \cdots & a_{mn}\end{bmatrix}\begin{bmatrix}d_1 & & & \\ & d_2 & & \\ & & \ddots & \\ & & & d_n\end{bmatrix}$$

$$=\begin{bmatrix}a_{11}d_1 & a_{12}d_2 & \cdots & a_{1n}d_n\\ a_{21}d_1 & a_{22}d_2 & \cdots & a_{2n}d_n\\ \cdots & \cdots & \cdots & \cdots\\ a_{m1}d_1 & a_{m2}d_2 & \cdots & a_{mn}d_n\end{bmatrix}。$$

(2)
$$\begin{bmatrix}d_1 & & & \\ & d_2 & & \\ & & \ddots & \\ & & & d_m\end{bmatrix}\begin{bmatrix}a_{11} & a_{12} & \cdots & a_{1n}\\ a_{21} & a_{22} & \cdots & a_{2n}\\ \cdots & \cdots & \cdots & \cdots\\ a_{m1} & a_{m2} & \cdots & a_{mn}\end{bmatrix}$$

$$=\begin{bmatrix}a_{11}d_1 & a_{12}d_1 & \cdots & a_{1n}d_1\\ a_{21}d_2 & a_{22}d_2 & \cdots & a_{2n}d_2\\ \cdots & \cdots & \cdots & \cdots\\ a_{m1}d_m & a_{m2}d_m & \cdots & a_{mn}d_m\end{bmatrix}$$

这由矩阵乘法定义不难证实。□

§7 系统的传递函数矩阵

设系统的 n 个状态变量为 $x_1(t),x_2(t),\cdots,x_n(t)$，这些变量在 $t=t_0$ 时的值都已知道，而且还知道 $t\geqslant t_0$ 时 r 个输入（或控制）变量

$u_1(t), u_2(t), \cdots, u_r(t)$，则在 $t \geq t_0$ 时系统的状态就完全确定。

系统的输出变量可以是某一个状态变量，但它们是完全不同的概念。状态变量是描述系统动态行为的信息，而输出则是人们希望从系统中获得的响应。

以 n 个状态变量为轴组成的空间叫做 n 维状态空间。而把上述 n 个状态变量看成这空间中一个向量 $x(t)$ 的 n 个分量，即

$$x(t) = (x_1(t), x_2(t), \cdots, x_n(t))^T。$$

$x(t)$ 称为系统的状态向量，它描述系统在 t 时刻的状态。

同样地，用

$$u(t) = (u_1(t), u_2(t), \cdots, u_r(t))^T$$

代表一个 r 维列向量。系统的状态方程是状态向量 $x(t)$ 的一阶微分方程：

$$\dot{x}(t) = f(x(t), u(t), t). \tag{1}$$

此式右边是状态向量 $x(t)$ 和输入向量 $u(t)$ 及时间 t 的向量函数。等式(1)也可以写成 n 个分量等式的形式。特别，对于线性定常系统，其状态方程为：

$$\dot{x}(t) = Ax(t) + Bu(t),$$

或更简单地写为：

$$\dot{x} = Ax + Bu. \tag{2}$$

这里 A, B 都是常数矩阵（元素都与 t 无关）：

$$A = (a_{ij})_{n \times n}, \quad B = (b_{ij})_{n \times r}。$$

其中 A 称为系统矩阵，B 称为输入（或控制）矩阵，二者都由系统本身的参数所组成。

类似地，线性定常系统的输出方程为

$$y = Cx + Du. \tag{3}$$

这里 C 是 $m \times n$ 矩阵（称为输出矩阵），D 是 $m \times r$ 矩阵（称为直接转移矩阵），$y = (y_1, y_2, \cdots, y_m)^T$ 是输出向量。矩阵 C 表达了输出变量与状态变量的关系，而 D 则表达了输入变量通过矩阵 D 所示关系

直接转移到输出。大多数实际系统中，矩阵 D 均为零矩阵。

当初始值为零，即 $x(0)=\theta$ 时，对(2)、(3)两式进行拉普拉斯变换(见第四章§6)可得

$$sX = AX + BU, \tag{4}$$

$$Y = CX + DU. \tag{5}$$

由(4)式得

$$(sE - A)X = BU.$$

在有理分式矩阵范围内，$sE-A$ 是可逆的，故有

$$X = (sE - A)^{-1}BU$$

代入(5)式得

$$Y = [C(sE - A)^{-1}B + D]U = GU \tag{6}$$

这里

$$G = C(sE - A)^{-1}B + D, \tag{7}$$

$G=G(s)$ 是 $m\times r$ 分式矩阵，称为系统的传递函数矩阵。将(6)式写成展开形式便是

$$\begin{bmatrix} Y_1(s) \\ Y_2(s) \\ \vdots \\ Y_m(s) \end{bmatrix} = \begin{bmatrix} G_{11}(s) & G_{12}(s) & \cdots & G_{1r}(s) \\ G_{21}(s) & G_{22}(s) & \cdots & G_{2r}(s) \\ \cdots & \cdots & \cdots & \cdots \\ G_{m1}(s) & G_{m2}(s) & \cdots & G_{mr}(s) \end{bmatrix} \begin{bmatrix} U_1(s) \\ U_2(s) \\ \vdots \\ U_r(s) \end{bmatrix}$$

传递函数矩阵 $G(s)$ 表达了输出向量 $Y(s)$ 与输入向量 $U(s)$ 之间的关系。它的每个元素 $G_{ij}(s)$ 表达了第 j 个输入 $U_j(s)$ 在第 i 个输出 $Y_i(s)$ 中的影响；或者说，第 i 个输出 $Y_i(s)$ 是全部 r 个输入 $U_j(s)$ $(j=1,2,\cdots,r)$ 通过各自的传递函数 $G_{i1},G_{i2},\cdots,G_{ir}$ 综合作用的结果。

例 1 设系统的状态方程及输出方程为：

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \\ \dot{x}_3 \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ -24 & -26 & -9 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 2 & -1 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}$$

$$\begin{bmatrix} y_1 \\ \\ y_2 \end{bmatrix} = \begin{bmatrix} 1 & -1 & 0 \\ \\ 0 & 1 & -1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$$

又初始值为零。求系统的传递函数矩阵 $G(s)$。

解 由于

$$(sE-A)^{-1} = \frac{\text{adj}(sE-A)}{|sE-A|}$$

$$= \frac{1}{s^3+9s^2+26s+24} \begin{bmatrix} s^2+9s+26 & s+9 & 1 \\ -24 & s^2+9s & s \\ -24s & -26s-24 & s^2 \end{bmatrix}$$

这里 $\text{adj}(sE-A)$ 为 $sE-A$ 的伴随矩阵。

因此由(7)式求得：

$$G(s) = \frac{1}{d}\begin{pmatrix} -s^2-7s+68 & s^2+6s-7 \\ 2s^2+46s+24 & -3s^2-33s-24 \end{pmatrix}$$

这里 $d=s^3+9s^2+26s+24$。

最后说明一下传递函数的意义。

若系统的传递函数 $G(s)=(a_{ij}(s)/b_{ij}(s))$ 满足条件

$$\deg(a_{ij}(s)) \leqslant \deg(b_{ij}(s)) \quad (\text{所有 } i,j),$$

则称 $G(s)$ 是真的；若此不等式为严格不等式($<$)，则称 $G(s)$ 是严格真的。可以证明这两种情形相当于下面的两个极限式子，即：

$G(s)$ 为真： $\lim\limits_{s\to\infty} G(s)=G_0$ （非零常数矩阵）；

$G(s)$ 严格真： $\lim\limits_{s\to\infty} G(s)=0$（矩阵）。

只有 $G(s)$ 是真的或严格真的，它所表征的系统才是可以用实际的物理元件来构成的，或才是能够正常地工作的。

§8 舒尔定理及矩阵的 QR 分解

舒尔(Schur)定理在理论上很重要，它是很多重要定理证明的出发点．而矩阵的 QR 分解在数值代数中起着重要作用，是计算矩阵特征值及求解线性方程组的一个重要工具．本节目的是介绍这两个结果．下面的讨论是在酉空间 C^n 内进行的．

定理 1(舒尔) 若 $A\in C^{n\times n}$，则存在酉矩阵 U，使得

$$U^H AU = T,$$

这里 T 为上三角矩阵，T 的(主)对角线上的元素都是 A 的特征值．

证明 设 A 的特征值为 $\lambda_1,\lambda_2,\cdots,\lambda_n$．若 ε_1 为 A 的属于 λ_1 的单位特征向量．把 ε_1 扩充成 C^n 的一组基：

$$\varepsilon_1,\eta_2,\cdots,\eta_n.$$

对它进行正交化、单位化，可以得到一组标准正交基

$$\varepsilon_1,\eta_2',\cdots,\eta_n'. \tag{1}$$

以这组基作列向量构成的矩阵

$$U_1 = (\varepsilon_1,\eta_2',\cdots,\eta_n')$$

为酉矩阵．由于

$$AU_1 = (A\varepsilon_1,A\eta_2',\cdots,A\eta_n') = (\lambda_1\varepsilon_1,A\eta_2',\cdots,A\eta_n'),$$

所以

$$U_1^H AU_1 = \begin{bmatrix} \bar{\varepsilon}_1^T \\ \bar{\eta}_2'^T \\ \vdots \\ \bar{\eta}_n'^T \end{bmatrix} (\lambda_1\varepsilon_1,A\eta_2',\cdots,A\eta_n').$$

注意到 $\bar{\varepsilon}_1^T\lambda\varepsilon_1=\lambda_1|\varepsilon_1|=\lambda_1$ 及向量组(1)的正交性，则有

$$U_1^H AU_1 = \begin{pmatrix} \lambda_1 & * \\ 0 & A_1 \end{pmatrix},$$

易知 $n-1$ 阶方阵 A_1 的特征值为 $\lambda_2,\cdots,\lambda_n$。设 $\varepsilon_2\in\in C^{n-1}$ 为 A_1 的属于 λ_2 的单位特征向量，又重复上述步骤，则又有 $n-1$ 阶酉矩阵 U_2，使得

$$U_2^H A_1 U_2=\begin{pmatrix}\lambda_2 & * \\ 0 & A_2\end{pmatrix}$$

令

$$V_2=\begin{bmatrix}1 & 0 \\ 0 & U_2\end{bmatrix}$$

则 V_2 和 U_1V_2 都是 n 阶酉矩阵，因而 $V_2^H U_1^H A U_1 V_2$ 具有如下形式

$$V_2^H U_1^H A U_1 V_2=\begin{bmatrix}\lambda_1 & & * \\ & \lambda_2 & \\ & & A_2\end{bmatrix}$$

继续这种作法，便得到 $n-i+1$ 阶的酉矩阵 $U_i(i=1,2,\cdots,n-1)$ 以及 n 阶酉矩阵 $V_i(i=2,3,\cdots,n-1)$。令

$$U=U_1V_2V_3\cdots V_{n-1},$$

则 U 为 n 阶酉矩阵，且 U^HAU 便给出了所要求的形式。又 $U^HAU=T$ 是上三角矩阵，显然 T 与 A 有相同的特征值。证毕。

注意：若 A 是实矩阵，且 A 的特征值恰好全为实数时，则特征向量可选为实向量，而上述步骤可以在实数运算下来完成，因而 U 可选为(实)正交矩阵。

因 U 为酉矩阵，$U^HU=E$，即 $U^H=U^{-1}$，故舒尔定理的结论亦可以叙述为：任一复数方阵都可以酉相似于上三角矩阵。

仿照这定理的证明可以得知，在定理的叙述中“上三角矩阵”改成“下三角矩阵”亦是可以的，当然它相应的酉矩阵与前者不同。又定理中的酉矩阵及上三角矩阵都不是唯一的。

应用定理 1 可以证明下面的定理。

定理 2 设 $A\in C^{n\times n}$，则有可逆矩阵 P，使得

$$P^{-1}AP = \begin{bmatrix} \lambda_1 & b_{12} & \cdots & b_{1n} \\ & \lambda_2 & & \\ & & \ddots & \vdots \\ 0 & & & \lambda_n \end{bmatrix}$$

而且$\sum\limits_{i,j=1}^{n} |b_{ij}| < \varepsilon$,其中 ε 是预先给定的任一正数。

证明 由定理 1,存在酉矩阵 Q,使得

$$Q^{-1}AQ = \begin{bmatrix} \lambda_1 & & & \rho_{ij} \\ & \lambda_2 & & \\ & & \ddots & \\ 0 & & & \lambda_n \end{bmatrix}$$

令 $F = \mathrm{diag}(r, r^2, \cdots, r^n)$,$r$ 为非零常数,且取 $T = QF$,则有

$$T^{-1}AT = F^{-1}\begin{bmatrix} \lambda_1 & & & \rho_{ij} \\ & \lambda_2 & & \\ & & \ddots & \\ 0 & & & \lambda_n \end{bmatrix}F$$

$$= \begin{bmatrix} \lambda_1 & r\rho_{12} & r^2\rho_{13} & \cdots & r^{n-1}\rho_{1n} \\ & \lambda_2 & r^2\rho_{23} & \cdots & r^{n-1}\rho_{2n} \\ & & \ddots & & \vdots \\ 0 & & & & \lambda_n \end{bmatrix}$$

$$= \begin{bmatrix} \lambda_1 & b_{12} & b_{13} & \cdots & b_{1n} \\ & \lambda_2 & b_{23} & \cdots & b_{2n} \\ & & \ddots & & \vdots \\ 0 & & & & \lambda_n \end{bmatrix}$$

对给定的 $\varepsilon > 0$,可选择 r,使得

$$\sum_{i,j=1}^{n} |b_{ij}| < \varepsilon$$

成立。证毕。

现在我们转到另一重要定理，它为计算特征值的数值方法提供了重要理论依据。

定理 3（QR 分解定理） 设 A 为 n 阶复数矩阵，则存在酉矩阵 Q 及上三角矩阵 R，使得

$$A = QR.$$

证明 把矩阵 A 写成分块矩阵形式

$$A = (a_1, a_2, \cdots, a_n),$$

其中 a_j 为 A 的第 j 个列向量，$j=1,2,\cdots,n$。注意，向量组 $a_1, a_2, \cdots, a_n$ 未必是线性无关的，亦即它可以不是酉空间 C^n 的一组基。现在我们从这组向量出发，用下述方法（类似于上章介绍过的施密特正交化过程），首先构作一组可以包含零向量在内的"正交向量组"（已讲过：零向量与任何向量正交）。

若 $a_1=\theta$，则取 $q_1=\theta$；若 $a_1\neq\theta$，就取

$$q_1 = \frac{a_1}{(a_1^H a_1)^{1/2}}。$$

接着作向量

$$y_2 = a_2 - (q_1^H a_2)q_1,$$

如果 $y_2=\theta$，就取 $q_2=\theta$；若 $y_2\neq\theta$，就取

$$q_2 = \frac{y_2}{(y_2^H y_2)^{1/2}}.$$

如此继续下去，求出 $q_1, q_2, \cdots, q_{k-1}$ 后，作

$$y_k = a_k - \sum_{i=1}^{k-1}(q_i^H a_k)q_i.$$

若 $y_k=\theta$，则取 $q_k=\theta$；若 $y_k\neq\theta$，就取

$$q_k = \frac{y_k}{(y_k^H y_k)^{1/2}} \quad (k = 2,3,\cdots,n)$$

可以验证 $q_1, y_2, y_3, \cdots, y_n$ 为"正交向量组"（它可以含有零向量），从而 $q_1, q_2, \cdots, q_n$ 是"正交向量组"，且每个向量 q_j 或为零向量，或为单位向量。而且每个 q_j 是 $a_1, a_2, \cdots, a_j$ 的线性组合。反过来上述

作法保证每个 a_j 是 $q_1,q_2,\cdots,q_j$ 的线性组合。因此存在复数 r_{kj} 使得

$$a_j = \sum_{k=1}^{j} r_{kj} q_k \quad (j = 1,2,\cdots,n) \tag{1}$$

若对所有 $k>j$，令 $r_{kj}=0$；而对所有使 $q_i=\theta$ 的每个 i，令 $r_{ij}=0$ $(j=1,2,\cdots,n)$，则经过上述过程，上三角矩阵 $R=(r_{ij})_{n\times n}$ 及向量 $q_1,q_2,\cdots,q_n$ 均由 $a_1,a_2,\cdots,a_n$ 所确定。矩阵 $Q=(q_1,q_2,\cdots,q_n)$ 的列是两两正交的，但有些列向量可能是零向量，不过(1)式表明这时已有 $A=QR$。但由于这样得到的 Q 可能仍不是酉矩阵，所以还需作如下处理。

取出 Q 中的所有非零向量 $q_{i_1},q_{i_2},\cdots,q_{i_s}$（标准正交组），并把它扩充成 C^n 的一组标准正交基：

$$q_{i_1},q_{i_2},\cdots,q_{i_s}, \quad z_1,z_2,\cdots,z_p,$$

数 p 当然恰为 Q 中零列向量的个数。现以 z_i 代替 Q 中第 i 个零列向量$(i=1,2,\cdots,p)$，用 Q' 表示这样替代后由 Q 产生的矩阵。因 Q' 的新列（即 Q 中经用 z_i 取代第 i 个零列所得到的列）与 R 中的零行（注意上述 $r_{ij}=0$ 的作法）相对应，于是 Q' 有标准正交列，且 $QR=Q'R$，故 $A=Q'R$ 便为所求的分解式。证毕。

这个定理的特殊情况是：当 A 为可逆矩阵时，则可取 R 为具有正对角元的上三角矩阵，并且这时的 Q,R 都是唯一的。此外，当 A 为实矩阵时，Q,R 都可取实矩阵。

§9 矩阵的奇异值分解

本节介绍的下述定理，称为矩阵的奇异值分解定理，它是在讨论最小二乘问题和广义逆矩阵计算以及很多应用领域中有着关键作用的一个定理。

定理 1 设 $A\in C^{m\times n}$，则存在酉矩阵 P,Q，使得

$$P^H AQ = \begin{bmatrix} D & 0 \\ 0 & 0 \end{bmatrix}$$

这里 $D=\mathrm{diag}(d_1,d_2,\cdots,d_r)$，且 $d_1 \geqslant d_2 \geqslant \cdots \geqslant d_r > 0$。

$d_i(i=1,2,\cdots,r)$称为 A 的奇异值，而

$$A = P\begin{bmatrix} D & 0 \\ 0 & 0 \end{bmatrix}Q^H$$

称为矩阵 A 的奇异值分解式。

证明 设 $B=A^H A$，则 $B^H=(A^H A)^H=A^H(A^H)^H=A^H A=B$，即 B 为 n 阶厄米特矩阵。可以证明(见注)B 的特征值均非负值。设这些特征值为 $d_1^2,d_2^2,\cdots,d_n^2$，其中 $d_1 \geqslant d_2 \geqslant \cdots \geqslant d_r > 0$，而其余的 d_i 均为零。设 $q_1,q_2,\cdots,q_n$ 为 B 的对应于 $d_1^2,d_2^2,\cdots,d_n^2$ 的标准正交特征向量。令

$$Q=(Q_1,Q_2),$$

其中 $Q_1=(q_1,q_2,\cdots,q_r)$， $Q_2=(q_{r+1},\cdots,q_n)$。

则有

$$Q_1^H A^H AQ_1 = D^2. \tag{1}$$

于是

$$D^{-1}Q_1^H A^H AQ_1 D^{-1} = E. \tag{2}$$

又由 Q_2 的定义可得

$$Q_2^H A^H AQ_2 = 0(\text{矩阵}),$$

从而

$$AQ_2 = 0(\text{矩阵}). \tag{3}$$

又令 $P_1=AQ_1D^{-1}$，则(2)式可写为

$$P_1^H P_1 = E. \tag{4}$$

(4)式表明 P_1 的列向量两两正交。又总可以用正交化过程选出 $m-r$ 个向量 $p_{r+1},p_{r+2},\cdots,p_m$，使它们和 P_1 的各个列向量构成 n 维酉空间 C^n 的一组标准正交基。又记 $P_2=(p_{r+1},p_{r+2},\cdots,p_m)$，则 $P=(P_1,P_2)$为一酉矩阵。

下面证明以上作出的 P,Q 即为定理中所需要的酉矩阵。因为

$$P^H AQ = \begin{pmatrix} P_1^H AQ_1 & P_1^H AQ_2 \\ P_2^H AQ_1 & P_2^H AQ_2 \end{pmatrix} \tag{5}$$

由(3)式可知：

$$P_1^H AQ_2 = P_2^H AQ_2 = 0(\text{矩阵}).$$

又由(1)式及 P_1 的定义可得

$$D^{-1}Q_1^H A^H AQ_1 = P_1^H AQ_1 = D.$$

而

$$P_2^H AQ_1 = P_2^H AQ_1 D^{-1} D = P_2^H P_1 D,$$

由 P_2 的选择方式可知 $P_2^H P_1 = 0$(矩阵)，所以有

$$P_2^H AQ_1 = 0(\text{矩阵}).$$

故(5)式化为：

$$P^H AQ = \begin{bmatrix} D & 0 \\ 0 & 0 \end{bmatrix}$$

证毕。

□注：设 x 为 $A^H A$ 的对应于特征值 λ 的特征向量，则由 $A^H Ax = \lambda x$，即有 $x^H A^H Ax = \lambda x^H x$，于是 $(Ax)^H (Ax) = \lambda x^H x$，即

$$(Ax, Ax) = \lambda(x, x).$$

但 $(x,x) > 0$，$(Ax, Ax) \geqslant 0$，所以 $\lambda \geqslant 0$。□

习 题 三

1. 若 $A^2 = E$，试证 A 的特征值只能是 ± 1。

2. 若 λ 是 A 的特征值，试从定义出发证明 λ^m 是 A^m 的特征值，这里 m 为正整数；又若 x 是 A 的属于特征值 λ 的特征向量，那么 A^m 的属于特征值 λ^m 的特征向量是什么？

3. 若 n 阶矩阵 A 的任意一行中 n 个元素的和都是 a，试证 $\lambda = a$ 是 A 的特征值，且 $x = (1,1,\cdots,1)^T$ 是 A 的对应于 λ 的特征向量。

4. 下列矩阵能否与对角形矩阵相似？若 A 能与对角形矩阵相似，则求出可逆矩阵 P，使得 $P^{-1}AP$ 为对角形矩阵：

(1) $A=\begin{pmatrix}3 & 4\\ 5 & 2\end{pmatrix}$

(2) $A=\begin{bmatrix}5 & -3 & 2\\ 6 & -4 & 4\\ 4 & -4 & 5\end{bmatrix}$

(3) $A=\begin{bmatrix}0 & 1 & 0\\ -4 & 4 & 0\\ -2 & 1 & 2\end{bmatrix}$

5. 若 A,B 均为 n 阶方阵，且有一个可逆，证明 AB 与 BA 相似，且有相同的特征多项式。

6. 在复数域上，求下列矩阵的约当标准形：

(1) $\begin{bmatrix}1 & -1 & 2\\ 3 & -3 & 6\\ 2 & -2 & 4\end{bmatrix}$　　(2) $\begin{bmatrix}3 & 7 & -3\\ -2 & -5 & 2\\ -4 & -10 & 3\end{bmatrix}$

(3) $\begin{bmatrix}3 & 0 & 8\\ 3 & -1 & 6\\ -2 & 0 & -5\end{bmatrix}$　　(4) $\begin{bmatrix}4 & 5 & -2\\ -2 & -2 & 1\\ -1 & -1 & 1\end{bmatrix}$

7. 证明：复数域上的任意 n 阶方阵 A，都存在可逆方阵 P，使得 $P^{-1}AP$ 为上三角矩阵。

8. 证明：(1) 方阵 A 的特征值全是零的充要条件，是存在自然数 m，使得 $A^m=0$(矩阵)；(2) 若 $A^m=0$(矩阵)，则 $|A+E|=1$。

9. 求下列多项式矩阵的史密斯标准形：

(1) $\begin{bmatrix}\lambda^3-\lambda & 2\lambda^2\\ \lambda^2+5\lambda & 3\lambda\end{bmatrix}$　　(2) $\begin{bmatrix}1-\lambda & \lambda^2 & \lambda\\ \lambda & \lambda & -\lambda\\ 1+\lambda^2 & \lambda^2 & -\lambda^2\end{bmatrix}$

(3) $\begin{bmatrix}\lambda^2+\lambda & 0 & 0\\ 0 & \lambda & 0\\ 0 & 0 & (\lambda+1)^2\end{bmatrix}$

并求出上述矩阵的不变因式及初级因子。

10. 利用特征多项式及哈密顿-开莱定理，证明：任意可逆矩阵 A 的逆

阵 A^{-1}都可以表示为 A 的多项式。

11. 设

$$A=\begin{pmatrix}1 & -1\\ 2 & 5\end{pmatrix}$$

证明：$B=2A^4-12A^3+19A^2-29A+37E$ 为可逆矩阵，并把 B^{-1}表示成 A 的多项式。

12. 若 A,B 均为 n 阶方阵，又 $E-AB$ 可逆，证明：

$$(E-BA)^{-1}=E+B(E-AB)^{-1}A.$$

13. 设

$$A=\begin{bmatrix}1 & 0 & 0\\ 1 & 0 & 1\\ 0 & 1 & 0\end{bmatrix}$$

证明：当 $n\geqslant 3$ 时，$A^n=A^{n-2}+A^2-E$，并求 A^{100}。

14. 若 A 满足 $A^2+A=2E$，证明 A 可与对角形矩阵相似。

15. 证明：任意方阵可表示为两个对称方阵的乘积，且其中一个是可逆的。

16. 判别下列多项式矩阵对是否为右互质的：

(1) $D(\lambda)=\begin{pmatrix}\lambda+1 & 0\\ \lambda^2+\lambda-2 & \lambda-1\end{pmatrix}$，$N(\lambda)=(\lambda+2,\lambda+1)$；

(2) $D(\lambda)=\begin{pmatrix}0 & -(\lambda+1)^2(\lambda+2)\\ (\lambda+2)^2 & \lambda+2\end{pmatrix}$，$N(\lambda)=\begin{pmatrix}\lambda & 0\\ -\lambda & \lambda^2\end{pmatrix}$

17. 判别下面的多项式矩阵 $M(\lambda)$是否为列既约和行既约：

$$M(\lambda)=\begin{bmatrix}\lambda^3+\lambda^2+1 & 2\lambda+1 & 2\lambda^2+\lambda+1\\ 2\lambda^3+\lambda-1 & 0 & 2\lambda^2+\lambda\\ 1 & \lambda-1 & \lambda^2-\lambda\end{bmatrix}$$

18. 求分式矩阵

$$G(\lambda)=\begin{bmatrix}\dfrac{\lambda}{(\lambda+1)^2(\lambda+2)^2} & \dfrac{\lambda}{(\lambda+2)^2}\\ \dfrac{-\lambda}{(\lambda+2)^2} & \dfrac{-\lambda}{(\lambda+2)^2}\end{bmatrix}$$

的史密斯-麦克米伦标准形及 $G(\lambda)$的一个右分解。

第四章　矩阵函数及其应用

本章在引入向量和矩阵的范数及极限的基础上，借助矩阵幂级数来定义矩阵函数，同时讨论了函数矩阵的微分、积分。以这些常被称为矩阵分析方面的基本知识为工具，将讨论线性系统中的几个问题。

§1　向量范数

在实内积空间及酉空间中，我们通过向量的内积定义了向量的长度$|x|=\sqrt{(x,x)}$。长度概念的引入已对问题的讨论扮演过重要的角色。对于一般的线性空间，我们可否引入一个类似长度而又比其含义更广泛的概念呢？所谓向量的范数便是这样的一个概念。

若V是实内积空间，$x,y\in V$为任意向量，a为实数域R中任一元素，则V中向量的长度具有下列三个基本性质：

(1) 当$x\neq\theta$时，都有$|x|>0$；

(2) $|ax|=|a|\cdot|x|$；

(3) $|x+y|\leqslant|x|+|y|$。

若V为酉空间，我们已经知道，只须将上面“a为实数域R中任一元素”改说成“a为复数域C中任一元素”，则V中向量的长度也具有上述三个性质。

特别地，对于欧氏空间R^n及酉空间C^n，其向量的长度当然也具有这三个性质，而且还可用向量的坐标更具体地表示出来。

现在我们考虑一般的线性空间。因为在其中未定义向量的内

积，所以不能按上述方式把向量的长度概念照搬进来。但我们希望在一般的线性空间里，也引入向量的某种度量概念，并使之保持上述的向量长度的三个基本性质。这样，我们就有下述向量范数的定义。

定义 1 设 V 是数域 P 上的线性空间。若对于 V 中任一向量 x，都有一非负实数 $\|x\|$ 与之对应，并且满足下列三个条件：

(1) 正定性：当 $x\neq\theta$ 时，都有 $\|x\|>0$；

(2) 齐次性：对于任何 $a\in P$，有 $\|ax\|=\|a\|\cdot\|x\|$；

(3) 三角不等式：对于任何 $x,y\in V$，都有：

$$\|x+y\|\leqslant\|x\|+\|y\|.$$

则称非负实数 $\|x\|$ 为向量 x 的范数。简言之，向量的范数是定义在线性空间上的非负实值函数。

这样，实内积空间及酉空间中向量的长度 $|x|=\sqrt{(x,x)}$ 都是向量的范数，因为 $|x|$ 满足定义 1 中 $\|x\|$ 的所有条件。

酉空间 C^n 的向量范数 $\|x\|$ 用向量长度 $|x|$ 来定义时，我们记作 $\|x\|_2$。亦即：

$$\|x\|_2=|x|=\sqrt{xx^H}=\sqrt{\sum_{i=1}^{n}\xi_i\bar{\xi}_i}=\sqrt{\sum_{i=1}^{n}|\xi_i|^2}$$

(对每个 $x=(\xi_1,\xi_2,\cdots,\xi_n)\in C^n$)。

当然，酉空间或实内积空间的向量范数并不都是前面所说的向量的长度。

例 1 若对酉空间 C^n 的每个向量 $x=(\xi_1,\xi_2,\cdots,\xi_n)$，定义

$$\|x\|_\infty=\max_{1\leqslant i\leqslant n}|\xi_i|$$

则易证它是 C^n 中的向量的范数。

事实上，当 $x\neq\theta$ 时，则 $\xi_1,\xi_2,\cdots,\xi_n$ 不全为零。故 $\|x\|_\infty=\max\limits_{1\leqslant i\leqslant n}|\xi_i|>0$，又对任意 $a\in C$，有

$$\|ax\|_\infty=\max_{1\leqslant i\leqslant n}|a\xi_i|$$

$$= |a| \cdot \max_{1 \leqslant i \leqslant n} |\xi_i|$$
$$= |a| \cdot \| x \|_\infty。$$

最后，我们来证明三角不等式也成立。若 $y = (\eta_1, \eta_2, \cdots, \eta_n)$ 为 C^n 中另一个任意向量，则

$$\| x + y \|_\infty = \max_{1 \leqslant i \leqslant n} |\xi_i + \eta_i|$$
$$\leqslant \max_{1 \leqslant i \leqslant n} |\xi_i| + \max_{1 \leqslant i \leqslant n} |\eta_i|$$
$$= \| x \|_\infty + \| y \|_\infty .$$

这就证明了 $\| x \|_\infty$ 是 C^n 中一种向量范数。

例 2 对于任意的 $x = (\xi_1, \xi_2, \cdots, \xi_n) \in C^n$，规定

$$\| x \|_1 = \sum_{i=1}^{n} |\xi_i|,$$

则 $\| x \|_1$ 也是空间 C^n 的向量范数。

证明 (1) 当 $x \neq \theta$ 时，则 $\xi_1, \xi_2, \cdots, \xi_n$ 不全为零，从而

$$\| x \|_1 = \sum_{i=1}^{n} |\xi_i| > 0;$$

(2) 对于任何 $a \in C$，则

$$\| ax \|_1 = \sum_{i=1}^{n} |a\xi_i| = \sum_{i=1}^{n} |a| |\xi_i| = |a| \cdot \| x \|_1;$$

(3) 若 $y = (\eta_1, \eta_2, \cdots, \eta_n) \in C^n$ 为任意向量，则

$$\| x + y \|_1 = \sum_{i=1}^{n} |\xi_i + \eta_i| \leqslant \sum_{i=1}^{n} (|\xi_i| + |\eta_i|)$$
$$= \| x \|_1 + \| y \|_1,$$

即三角不等式也成立。

一般地，可以证明：若定义

$$\| x \|_p = \left(\sum_{i=1}^{n} |\xi_i|^p \right)^{\frac{1}{p}} \quad (1 \leqslant p < \infty),$$

则 $\| x \|_p$ 也是酉空间 C^n 的向量范数，称为向量 x 的 p-范数。关于这一点，证明是相当繁复的，在此就不作介绍了。

显然，当 $p=1$ 时，即为例 2 的范数 $\|x\|_1$；当 $p=2$ 时，即得到上面讲过的范数 $\|x\|_2=\sqrt{\sum_{i=1}^{n}|\xi_i|^2}$。现在我们来证明：

$$\|x\|_\infty=\lim_{p\to\infty}\|x\|_p=\max_{1\leqslant i\leqslant n}|\xi_i| \quad \text{（例 1 的范数）。}$$

证明 当 $x=\theta$，上式显然成立。故只需对非零向量 x 加以证明。令 $\omega=\max\limits_{1\leqslant i\leqslant n}|\xi_i|$，则有

$$\begin{aligned}\left(\sum_{i=1}^{n}|\xi_i|^p\right)^{\frac{1}{p}}&=\left(\sum_{i=1}^{n}\omega^p\left|\frac{\xi_i}{\omega}\right|^p\right)^{\frac{1}{p}}\\&=\omega\left(\sum_{i=1}^{n}\left|\frac{\xi_i}{\omega}\right|^p\right)^{\frac{1}{p}}\\&=\omega\left(\sum_{i=1}^{n}\beta_i^p\right)^{\frac{1}{p}},\end{aligned}$$

这里 $\beta_i=\left|\frac{\xi_i}{\omega}\right|\leqslant 1$，又至少有一个 $\beta_i=1$，所以有

$$1\leqslant\sum_{i=1}^{n}\beta_i^p\leqslant n。$$

因此，

$$1\leqslant\left(\sum_{i=1}^{n}\beta_i^p\right)^{\frac{1}{p}}\leqslant n^{\frac{1}{p}}.$$

又因为 $\lim\limits_{p\to\infty}n^{\frac{1}{p}}=1$，所以

$$\lim_{p\to\infty}\left(\sum_{i=1}^{n}\beta_i^p\right)^{\frac{1}{p}}=1.$$

从而

$$\lim_{p\to\infty}\left(\sum_{i=1}^{n}|\xi_i|^p\right)^{\frac{1}{p}}=\omega,$$

即

$$\|x\|_\infty=\max_{1\leqslant i\leqslant n}|\xi_i|.$$

由前面的讨论可见，有限维线性空间（如 R^n 及 C^n）上可以引入

各种各样的向量范数。范数的种数可无穷多，但这些范数之间有重要的关系。

定理 1 对于任何有限维向量空间 V 上定义的任意两种向量范数 $\|x\|_a$ 及 $\|x\|_b$，都存在两个与 x 无关的正的常数 C_1, C_2，使得对 V 中任一向量 x，都有

$$\|x\|_a \leqslant C_1 \|x\|_b, \quad \|x\|_b \leqslant C_2 \|x\|_a. \qquad (1)$$

满足(1)的两个不等式的两种向量范数称为等价的。因此，定理 1 亦可叙述为：有限维向量空间上的不同向量范数是等价的。

证明 为简单起见，我们仅就实数域 R 上的 n 维线性空间 V 来证明这一定理。其实对于复空间 V（如酉空间 C^n），证明也是类似的。

设 $e_1, e_2, \cdots, e_n$ 是 V 的一组基，于是 V 中任意向量 x 可以表示为

$$x = \xi_1 e_1 + \xi_2 e_2 + \cdots + \xi_n e_n.$$

定义

$$\|x\|_E = \sqrt{\xi_1^2 + \xi_2^2 + \cdots + \xi_n^2}.$$

由前所述，它显然是一种向量范数。又给定的范数 $\|x\|_a$ 具有形式

$$\|x\|_a = \|\xi_1 e_1 + \xi_2 e_2 + \cdots + \xi_n e_n\|_a.$$

现在我们来证明 $\|x\|_a$ 与 $\|x\|_E$ 等价。

上述的 $\|x\|_a$ 可以看作 n 个变量 $\xi_1, \xi_2, \cdots, \xi_n$ 的函数，记

$$\varphi(\xi_1, \xi_2, \cdots, \xi_n) = \|x\|_a.$$

现在证明 $\varphi(\xi_1, \xi_2, \cdots, \xi_n)$ 是连续函数。

设另一向量为

$$x' = \xi_1' e_1 + \xi_2' e_2 + \cdots + \xi_n' e_n,$$

其范数为

$$\|x'\|_a = \varphi(\xi_1', \xi_2', \cdots, \xi_n').$$

则有

$$|\varphi(\xi_1', \xi_2', \cdots, \xi_n') - \varphi(\xi_1, \xi_2, \cdots, \xi_n)|$$

$$=|\|x'\|_a-\|x\|_a|$$
$$\leqslant\|x'-x\|_a \quad (\text{见习题四})$$
$$=\|(\xi_1'-\xi_1)e_1+(\xi_2'-\xi_2)e_2+\cdots+(\xi_n'-\xi_n)e_n\|_a$$
$$\leqslant|\xi_1'-\xi_1|\cdot\|e_1\|_a+|\xi_2'-\xi_2|\cdot\|e_2\|_a+\cdots$$
$$+|\xi_n'-\xi_n|\cdot\|e_n\|_a.$$

因为 $\|e_i\|_a(i=1,2,\cdots,n)$ 是常数，因此当 ξ_i' 与 ξ_i 充分接近时，$\varphi(\xi_1',\xi_2',\cdots,\xi_n')$ 就充分接近 $\varphi(\xi_1,\xi_2,\cdots,\xi_n)$，即 $\varphi(\xi_1,\xi_2,\cdots,\xi_n)$ 是一个连续函数。

根据连续函数的性质，可知在有界闭集

$$W=\{x=(\xi_1,\xi_2,\cdots,\xi_n)\mid\xi_1^2+\xi_2^2+\cdots+\xi_n^2=1\}$$

上，函数 $\|x\|_a=\varphi(\xi_1,\xi_2,\cdots,\xi_n)$ 可达到最大值 M 及最小值 m。而当 $x\in W$ 时，显然 $x\neq\theta$。因此有 $m>0$。又记

$$d=\sqrt{\sum_{i=1}^{n}\xi_i^2} \quad (x=(\xi_1,\xi_2,\cdots,\xi_n)\in V).$$

则向量

$$y=\sum_{i=1}^{n}\frac{\xi_i}{d}e_i=\frac{1}{d}x$$

的分量满足

$$\sum_{i=1}^{n}\left(\frac{\xi_i}{d}\right)^2=1,$$

因此 $y\in W$；于是

$$0<m\leqslant\|y\|_a=\varphi\left(\frac{\xi_1}{d},\frac{\xi_2}{d},\cdots,\frac{\xi_n}{d}\right)\leqslant M.$$

由 $x=\mathrm{d}y$ 得

$$\|x\|_a=\|\mathrm{d}y\|_a=|\mathrm{d}|\cdot\|y\|_a=\mathrm{d}\cdot\|y\|_a.$$

再由上式即得

$$md\leqslant\|x\|_a\leqslant Md;$$

即

$$m\|x\|_E\leqslant\|x\|_a\leqslant M\|x\|_E.$$

若取 $C_1=M_a C_2=1/m$，则由此不等式即得：

$$\|x\|_a \leqslant C_1\|x\|_E, \quad \|x\|_E \leqslant C_2\|x\|_a.$$

这就证明了 $\|x\|_a$ 与 $\|x\|_E$ 是等价的。

同样可以证明 $\|x\|_b$ 与 $\|x\|_E$ 等价：

$$\|x\|_b \leqslant C_3\|x\|_E, \quad \|x\|_E \leqslant C_4\|x\|_b,$$

因而有

$$\|x\|_a \leqslant C_1C_4\|x\|_b, \quad \|x\|_b \leqslant C_3C_2\|x\|_a.$$

即 $\|x\|_a$ 与 $\|x\|_b$ 等价。定理证毕。

同一个向量 $x\in C^n$，按不同公式所定义的向量范数，其大小一般不相等。例如，当取 $x=(1,1,\cdots,1)\in C^n$，则：

$$\|x\|_2=\sqrt{n}, \|x\|_1=n, \|x\|_\infty=1.$$

但在考虑向量序列的收敛性时，则效果是一致的，即此序列在一种范数定义下收敛，则在另一种范数定义下亦收敛，且极限相同。这也就是所谓向量范数的等价性，上面的定理 1 就用不等式刻划了这一性质。注意：在无限维线性空间中，两个向量范数是可以不等价的。

§2 矩阵范数

一个 $m\times n$ 矩阵也可以看作是一个 mn 维向量，因此可以按向量范数的办法来定义范数。但矩阵还有矩阵之间的乘法运算，因此，对于 $n\times n$ 的方阵 A，我们定义它的范数如下。

定义 1 在 $P^{n\times n}$ 上定义一个非负实值函数 $\|A\|$（对每个 $A\in P^{n\times n}$），如果对任意的 $A,B\in P^{n\times n}$ 都满足下列四个条件：

(1) 正定性：若 $A\neq 0$（矩阵），则 $\|A\|>0$；

(2) 齐次性：对任意 $a\in P$，有 $\|aA\|=|a|\cdot\|A\|$；

(3) 三角不等式：$\|A+B\|\leqslant\|A\|+\|B\|$；

(4) $\|AB\| \leqslant \|A\| \cdot \|B\|$.

则非负实函数 $\|A\|$ 称为 $n\times n$ 方阵的范数。

与向量的情形一样，矩阵也可以有各种各样的范数，而且在大多数情况下，矩阵范数常和向量范数混合在一起使用，因此我们考虑的一些矩阵范数，应当使它们与向量范数联系起来。下面的概念就反映这种联系。

定义 2 若对任何 $A\in P^{n\times n}$ 及 n 维列向量 $x\in P^n$，方阵范数 $\|A\|$ 能与某种向量范数 $\|x\|_a$ 满足关系式

$$\|Ax\|_a \leqslant \|A\| \cdot \|x\|_a,$$

则称方阵范数 $\|A\|$ 与向量范数 $\|x\|_a$ 是相容的。

可以证明：

1) $P^{n\times n}$ 上的每一种方阵范数，在 P^n 上都存在与它相容的向量范数；

2) $P^{n\times n}$ 上任意两种方阵范数 $\|A\|_\alpha$，$\|A\|_\beta$ 都是等价的，即存在两个与 A 无关的正数 C_1, C_2，使得

$\|A\|_\alpha \leqslant C_1\|A\|_\beta$，$\|A\|_\beta \leqslant C_2\|A\|_\alpha \quad (\forall\ A\in P^{n\times n})$；

3) 若 $A=(a_{ij})_{n\times n}\in C^{n\times n}$，则

$$\|A\|_F = \sqrt{\sum_{i,j=1}^{n}|a_{ij}|^2} = \sqrt{\operatorname{tr}(A^H A)}$$

是一种与向量范数 $\|x\|_2=\sqrt{\sum_{i=1}^{n}|\xi_i|^2}$ 相容的方阵范数，称为 Frobenius 范数（$x\in C^n$）。

前两点的证明从略。兹证 3) 如下：

(1) 当 $A\neq 0$（矩阵），则 $\|A\|_F>0$（矩阵）显然成立；

(2) 对任意 $k\in C$，则

$$\|kA\|_F = \sqrt{\sum_{i,j=1}^{n}|ka_{ij}|^2}$$

$$=\sqrt{|k|^2\sum_{i,j=1}^{n}|a_{ij}|^2}$$

$$=|k|\cdot\|A\|_F;$$

(3) $$\|A+B\|_F=\sqrt{\sum_{i,j=1}^{n}|a_{ij}+b_{ij}|^2}$$

$$\leqslant\sqrt{\sum_{i,j=1}^{n}(|a_{ij}|^2+|b_{ij}|^2+2|a_{ij}|\cdot|b_{ij}|)}$$

$$\leqslant\sqrt{\|A\|_F^2+\|B\|_F^2+2\sqrt{\sum_{i,j=1}^{n}|a_{ij}|^2}\cdot\sqrt{\sum_{i,j=1}^{n}|b_{ij}|^2}}$$

$$=\sqrt{(\|A\|_F+\|B\|_F)^2}=\|A\|_F+\|B\|_F;$$

(4) $$\|AB\|_F=\sqrt{\sum_{i,j=1}^{n}|a_{i1}b_{1j}+a_{i2}b_{2j}+\cdots+a_{in}b_{nj}|^2}$$

$$\leqslant\sqrt{\sum_{i,j=1}^{n}(|a_{i1}|^2+\cdots+|a_{in}|^2)(|b_{1j}|^2+\cdots+|b_{nj}|^2)}\ ^{*}$$

$$=\sqrt{\sum_{i,j=1}^{n}|a_{ij}|^2}\cdot\sqrt{\sum_{i,j=1}^{n}|b_{ij}|^2}$$

$$=\|A\|_F\cdot\|B\|_F;$$

(5) 设 $x=(\xi_1,\xi_2,\cdots,\xi_n)^T\in C^n$，$A=(a_{ij})_{n\times n}\in C^{n\times n}$，令

$$Ax=\begin{pmatrix}\eta_1\\ \eta_2\\ \vdots\\ \eta_n\end{pmatrix}=y,$$

---

* 注：此处应用不等式$(\sum_{i=1}^{m}a_ib_i)^2\leqslant(\sum_{i=1}^{m}a_i^2)(\sum_{i=1}^{m}b_i^2)$，这里 a_i,b_i 为非负实数，$i=1,2,\cdots,m$）

则有

$$\| Ax \|_2 = \| y \|_2 = \sqrt{\sum_{k=1}^{n} |\eta_k|^2}$$

$$= \sqrt{\sum_{k=1}^{n} |a_{k1}\xi_1 + a_{k2}\xi_2 + \cdots + a_{kn}\xi_n|^2}$$

$$\leqslant \sqrt{\sum_{k=1}^{n} (|a_{k1}| \cdot |\xi_1| + \cdots + |a_{kn}| \cdot |\xi_n|)^2}$$

$$\leqslant \sqrt{\sum_{k=1}^{n} (\sum_{j=1}^{n} |a_{kj}|^2)(\sum_{j=1}^{n} |\xi_j|^2)}$$

$$= \sqrt{\sum_{k,j=1}^{n} |a_{kj}|^2} \cdot \sqrt{\sum_{j=1}^{n} |\xi_j|^2}$$

$$= \| A \|_F \cdot \| x \|_2.$$

即 $\| A \|_F$ 是与向量范数 $\| x \|_2$ 相容的矩阵范数。证毕.

F-范数的优点之一是乘以酉矩阵 U 后不变(在实矩阵的情形下是乘以正交矩阵后不变),即

$$\| UA \|_F = \| A \|_F = \| AU \|_F.$$

事实上,

$$\| UA \|_F^2 = \mathrm{tr}[(UA)^H(UA)]$$

$$= \mathrm{tr}[A^H(U^HU)A]$$

$$= \mathrm{tr}[A^HA] = \| A \|_F^2,$$

即 $\qquad \| UA \|_F = \| A \|_F.$

又因为 $\| A \|_F = \| A^H \|_F$,且 U^H 也是酉矩阵,故由上面的结果即得:

$$\| AU \|_F = \| (AU)^H \|_F = \| U^HA^H \|_F = \| A^H \|_F = \| A \|_F.$$

由此可知,A 的酉相似矩阵的 F-范数是相同的,即:

若 $B = U^HAU$,则 $\| B \|_F = \| A \|_F$.

除 F-范数外,下列三种也是比较常用的矩阵范数.

1） $\| A \|_1 = \max\limits_{1\leqslant j\leqslant n} \sum\limits_{i=1}^{n} |a_{ij}|$ （列模和最大者）。

2） $\| A \|_\infty = \max\limits_{1\leqslant i\leqslant n} \sum\limits_{j=1}^{n} |a_{ij}|$ （行模和最大者）。

3） $\| A \|_2 = \sqrt{\lambda_{A^HA}}$ （λ_{A^HA}是 A^HA 的最大特征值）。

□注：因$(A^HA)^H = A^H(A^H)^H = A^HA$，即 A^HA 是 Hermite 矩阵，它对应的二次型

$$f(x) = x^H(A^HA)x = (Ax)^H(Ax) = y^Hy \geqslant 0$$

是正定或半正定的，因此它的特征值都大于或等于零。□

到此为止，我们已把向量和矩阵的范数作了一个初步的讨论。这主要是为本书后面的一些章节服务的，同时也为读者进一步的学习和应用打下基础。在数值分析及各种优化问题的讨论中，矩阵或向量的范数都有重要的应用。例如在摄动理论中，需要考虑当方阵 A 有摄动（微小变化）而变成 $A+\varepsilon$ 时，方程 $Ax=b$ 的解有何变化？A 的逆阵及 A 的特征值又怎样变化（A 为可逆方阵）？这些问题的解决都有赖于向量和矩阵的范数。欲知其详，可参考书末所列的参考书。

§3 向量和矩阵的极限

定义 1 若 $x^{(m)} = (\xi_1^{(m)}, \xi_2^{(m)}, \cdots, \xi_n^{(m)}) \in C^n (m=1,2,\cdots)$，如果存在极限

$$\lim_{m\to\infty} \xi_i^{(m)} = \xi_i \quad (i = 1,2,\cdots,n),$$

则称酉空间 C^n 的向量序列$\{x^{(m)}\}$收敛于向量 $x=(\xi_1,\xi_2,\cdots,\xi_n)$，并记为

$$\lim_{m\to\infty} x^{(m)} = x \text{ 或 } x^{(m)} \to x.$$

换言之，向量序列的极限是按坐标序列的极限来定义的。当向量序

列不收敛时，也叫做发散的。

定理 1

$$\lim_{m\to\infty} x^{(m)}=x \Leftrightarrow \lim_{m\to\infty}\|x^{(m)}-x\|=0$$ （对任一向量范数 $\|\cdot\|$）。

证明 利用向量范数的等价性，易知只要对一种向量范数进行证明，则对任何一种向量范数也能成立。为此，取向量范数 $\|x\|_\infty$。

如果对向量范数 $\|x\|_\infty$，有

$$\lim_{m\to\infty}\|x^{(m)}-x\|=0,$$

则由

$$\|x^{(m)}-x\|_\infty=\max_{1\leq i\leq n}|\xi_i^{(m)}-\xi_i|\to 0 \quad (m\to\infty)$$

可知对每个 i $(i=1,2,\cdots,n)$ 有 $\xi_i^{(m)}\to\xi_i$。因此，

$$\lim_{m\to\infty} x^{(m)}=x。$$

反之，若有 $\lim\limits_{m\to\infty} x^{(m)}=x$，则由定义，这相当于

$$\xi_i^{(m)}-\xi_i\to 0 \quad (i=1,2,\cdots,n)。$$

故对任给正数 ε，都有正数 M_i 存在，使得 $m>M_i$ 时，都有

$$|\xi_i^{(m)}-\xi_i|<\varepsilon \quad (i=1,2,\cdots,n)。$$

若取 $M=\max\limits_{1\leq i\leq n}\{M_i\}$，则当 $m>M$ 时，对每个 i 值，上述不等式都成立，从而，$m>M$ 时，

$$\|x^{(m)}-x\|_\infty=\max_{1\leq i\leq n}|\xi_i^{(m)}-\xi_i|<\varepsilon。$$

这就证明了

$$\lim_{m\to\infty}\|x^{(m)}-x\|_\infty=0。$$

定理证毕。

这定理表明：向量序列 $\{x^{(m)}\}$ 收敛于向量 x，当且只当对任何一种向量范数 $\|x\|$，序列 $\{\|x^{(m)}-x\|\}$ 收敛于零。因此 n 维向量序列的收敛问题，借助范数概念为工具，就归结为实数序列的收敛

问题。在研究高维空间的收敛性时，不仅向量范数具有这个优点，矩阵范数亦如此。

定义 2 若 $A_m=(a_{ij}^{(m)})\in C^{n\times n}(m=1,2,\cdots)$，如果存在极限

$$\lim_{m\to\infty} a_{ij}^{(m)}=a_{ij}\quad(i,j=1,2,\cdots,n),$$

则称方阵序列$\{A_m\}$收敛于方阵 $A=(a_{ij})\in C^{n\times n}$，记为

$$\lim_{m\to\infty} A_m=A，\text{或 } A_m\to A\quad(m\to\infty).$$

当方阵序列不收敛时，也称为发散的。

由此可见，方阵序列$\{A_m\}$的收敛性，相当于 n^2 个数值序列$\{a_{ij}^{(m)}\}$ $(i,j=1,2,\cdots,n)$的收敛性。

例如，若

$$A_m=\begin{pmatrix}\dfrac{1}{m} & \dfrac{\sqrt{2}m-1}{3m+2}\\ 1-\dfrac{1}{m^2} & \sqrt{-1}\cos\dfrac{\pi}{m}\end{pmatrix}\in C^{2\times 2}\quad m=1,2,\cdots,$$

则有

$$\lim_{m\to\infty} A_m=\begin{pmatrix}0 & \dfrac{\sqrt{2}}{3}\\ 1 & \sqrt{-1}\end{pmatrix}$$

与定理 1 类似，我们有

定理 2

$$\lim_{m\to\infty} A_m=A\Leftrightarrow\lim_{m\to\infty}\|A_m-A\|=0$$ （对任一方阵范数 $\|\cdot\|$）。

证明 由方阵范数的等价性，我们只需对 F-范数进行证明。我们有：

$$\lim_{m\to\infty} A_m=A\Leftrightarrow\lim_{m\to\infty}(a_{ij}^{(m)}-a_{ij})=0\quad(\forall\ i,j)$$

$$\Leftrightarrow\lim_{m\to\infty}\sqrt{\sum_{i,j=1}^{n}|a_{ij}^{(m)}-a_{ij}|^2}\left(=\lim_{m\to\infty}\|A_m-A\|_F\right)=0.$$

这就是所要证明的。

定理 2 表明:方阵序列$\{A_m\}$收敛于方阵 A,当且只当对任一方阵范数$\|\cdot\|$,序列$\|A_m-A\|$收敛于零。特别地,$A_m\to 0$(矩阵),当且只当$\|A_m\|\to 0\ (m\to\infty)$。

$C^{n\times n}$中收敛的方阵序列有下列基本性质:

(1) 若$\lim\limits_{m\to\infty}A_m=A$,则对$C^{n\times n}$中任何方阵范数$\|\cdot\|$,$\|A_m\|$有界。

证明 由于方阵范数的等价性,我们只需在 F-范数下作出证明。由于$A_m\to A$,故由定理 2,我们有$\|A_m-A\|_F\to 0$(矩阵),所以数列$\|A_m-A\|_F$有界. 故有非负实数C_1,使得

$$\|A_m-A\|_F\leqslant C_1,$$

从而有

$$\begin{aligned}\|A_m\|_F&=\|A_m-A+A\|_F\leqslant\|A_m-A\|_F+\|A\|_F\\&\leqslant C_1+\|A\|_F=C_1+C_2=C,\end{aligned}$$

即$\|A_m\|_F$是有界的。(这里 A 是给定矩阵,$\|A\|_F=C_2$也是常数)。

(2) 若$A_m\to A$,$B_m\to B$,又$a_m\to a$,$b_m\to b$(这里$\{a_m\}$,$\{b_m\}$为数列),则有

$$\lim_{m\to\infty}(a_mA_m+b_mB_m)=aA+bB,$$

$$\lim_{m\to\infty}A_mB_m=AB.$$

证明 这两个式子的证明是类似的,我们只给出后一等式的证明。由于在任何一种取定的方阵的范数$\|\cdot\|$下,

$$\begin{aligned}\|A_mB_m-AB\|&=\|A_mB_m-A_mB+A_mB-AB\|\\&\leqslant\|A_mB_m-A_mB\|+\|A_mB-AB\|\\&\leqslant\|A_m\|\cdot\|B_m-B\|+\|A_m-A\|\cdot\|B\|,\end{aligned}$$

但由(1)知$\|A_m\|$有界,又由定理 2 知$\|A_m-A\|\to 0$,$\|B_m-B\|\to 0$,而$\|B\|$是确定常数,当然是有界的,故由上面的不等式及这些

说明，即有：

对任给正数 ε，当 $m>N$，$\| A_mB_m - AB \| < \varepsilon$。即 $\lim\limits_{m\to\infty} \| A_mB_m - AB \| = 0$，故 $\lim\limits_{m\to\infty} A_mB_m = AB$。

(3) 若 $A_m \to A$，且 A_m^{-1} 及 A^{-1} 都存在，则

$$\lim_{m\to\infty} A_m^{-1} = A^{-1}.$$

证明 对二阶方阵，易知

$$\lim_{m\to\infty} |A_m| = |A|$$

成立；应用归纳法且按一行（列）展开，则可证明此式对 n 阶方阵也成立。再注意到伴随矩阵 A_m^* 的元素都是 $n-1$ 阶行列式，因此有

$$\lim_{m\to\infty} A_m^{-1} = \lim_{m\to\infty} \frac{1}{|A_m|} A_m^* = \frac{1}{|A|} A^* = A^{-1}.$$

下面我们来考察由方阵 $A \in C^{n\times n}$ 的幂所形成的序列 $\{A^m\}$：$E, A, A^2, \cdots, A^m, A^{m+1}, \cdots$，研究其收敛于零矩阵的条件。首先我们有

定理 3 $\lim\limits_{m\to\infty} A^m = 0$（矩阵）的充分条件，是有某一方阵范数 $\| \cdot \|$，使得 $\| A \| < 1$。

证明 由方阵范数定义的条件(4)，即有

$$\| A^m \| \leqslant \| A^{m-1} \| \cdot \| A \| \leqslant \| A^{m-2} \| \cdot \| A \|^2 \leqslant \cdots \leqslant \| A \|^m,$$

因此，若 $\| A \| < 1$，则 $\| A \|^m \to 0$，从而 $\| A^m \| \to 0$。由定理 2 便得 $\lim\limits_{m\to\infty} A^m = 0$（矩阵）。证毕。

方阵序列 $\{A^m\}$ 收敛于零矩阵与方阵的特征值有下面重要联系。

定理 4 $\lim\limits_{m\to\infty} A^m = 0$（矩阵）的充分必要条件，是 A 的所有特征值的模都小于 1。

证明 设 A 的约当标准形为

$$J=\begin{pmatrix} J_{r_1}(\lambda_1) & & & \\ & J_{r_2}(\lambda_2) & & \\ & & \ddots & \\ & & & J_{r_p}(\lambda_p) \end{pmatrix}$$

这里

$$J_{r_i}(\lambda_i)=\begin{pmatrix} \lambda_i & & & \\ 1 & \lambda_i & & \\ & \ddots & \ddots & \\ & & 1 & \lambda_i \end{pmatrix}_{r_i\times r_i}$$

由于 $A=TJT^{-1}$，所以 $A^m=TJ^mT^{-1}$。而且

$$J^m=\begin{pmatrix} J^m_{r_1}(\lambda_1) & & & \\ & J^m_{r_2}(\lambda_2) & & \\ & & \ddots & \\ & & & J^m_{r_p}(\lambda_p) \end{pmatrix}$$

但是不难证明(见下面的注释)：

$$J^m_{r_i}(\lambda_i)=\begin{pmatrix} f_m(\lambda_i) & & & \\ f'_m(\lambda_i) & f_m(\lambda_i) & & \\ \vdots & \ddots & \ddots & \\ \dfrac{f_m^{(r_i-1)}(\lambda_i)}{(r_i-1)!} & \cdots\cdots & f'_m(\lambda_i) & f_m(\lambda_i) \end{pmatrix} \quad (*)$$

这里 $f_m(\lambda)=\lambda^m$，又 $f_m(\lambda)$在 $\lambda=\lambda_i$ 时的 t 阶导数为：

$$f_m^{(t)}(\lambda_i)=m(m-1)\cdots(m-t+1)\lambda_i^{m-t}$$

$$=\frac{m!\,\lambda_i^{m-t}}{(m-t)!} \quad (t=0,1,2,\cdots,r_i-1).$$

由此可以看出，当 $m\to\infty$时，下列各个陈述的等价性：

$$A^m\to 0 \Leftrightarrow J^m\to 0 \Leftrightarrow \text{每个 } J^m_{r_i}(\lambda_i)\to 0(\text{矩阵})$$

$\Leftrightarrow$ 对每个 λ_i 及每个 t 值，$f_m^{(t)}(\lambda_i) \to 0 \Leftrightarrow |\lambda_i| < 1$.

$(i = 1,2,\cdots,p)$

这就证明了定理 4.

□注：若

$$H_n = \begin{bmatrix} 0 & & & \\ 1 & 0 & & \\ & \ddots & \ddots & \\ & & 1 & 0 \end{bmatrix}_{n\times n}$$

则由归纳法可证明：当 $p\geqslant n$ 时，都有 $H_n^p=0$(矩阵)。我们指出，n 阶矩阵 H_n 具有这样一个性质：每自乘一次，就恰好把 H_n 中那条"全1线"往左下角方向平移一位(1 的个数减少了一个)，自乘 $n-1$ 次，亦即在 H_n^{n-1} 中，"全 1 线"已退化为左下角单独一个元素 1，而 H_n^{n-1} 中的其他元素均为 0。再用 H_n 去乘 H_n^{n-1}，又相当于把 H_n^{n-1} 中那个 1 从左下角方向移走，而以 0 代替它。于是 $H_n^n=0$(矩阵)。自然有 $H_n^{n+1}=H_n^{n+2}=\cdots=0$(矩阵)。

又显然每个 r_i 阶约当块 $J_{r_i}(\lambda_i)$可以写成如下形式

$$J_{r_i}(\lambda_i) = H_{r_i} + \lambda_i E,$$

其中 H_{r_i}是上述形式的 r_i 阶方阵。E 是 r_i 阶单位方阵。由于 H_i 与 $\lambda_i E$ 具有交换性，即 $H_i \cdot (\lambda_i E) = (\lambda_i E) \cdot H_i$，故注意到 $m' \geqslant r_i$ 时，$H_{r_i}^{m'}=0$(矩阵)，所以

$$\begin{aligned} J_{r_i}^m(\lambda_i) &= (H_{r_i}+\lambda_i E)^m \\ &= \lambda_i^m E + C_m^1 \lambda_i^{m-1} E \cdot H_{r_i} + \cdots + C_m^{r_i-1} \lambda_i^{m-r_i+1} E \cdot H_{r_i}^{r_i-1} \end{aligned}$$

$$= \begin{bmatrix} \lambda_i^m & & & \\ & \lambda_i^m & & \\ & & \ddots & \\ & & & \lambda_i^m \end{bmatrix} + \begin{bmatrix} 0 & & & \\ m\lambda_i^{m-1} & 0 & & \\ & \ddots & \ddots & \\ & & m\lambda_i^{m-1} & 0 \end{bmatrix}$$

$$
+\begin{pmatrix}
0 & & & & \\
0 & 0 & & & \\
\frac{m(m-1)}{2!}\lambda_i^{m-2} & & \ddots & & \\
 & \ddots & & \ddots & \\
 & & \frac{m(m-1)}{2!}\lambda_i^{m-2} & 0 & 0
\end{pmatrix}+\cdots
$$

$$
+\begin{pmatrix}
0 & \cdots & 0 \\
\vdots & & \vdots \\
0 & & \vdots \\
C_m^{r_i-1}\lambda_i^{m-r_i+1} & 0\cdots & 0
\end{pmatrix} \quad (m \geqslant r_i - 1)
$$

若令 $\lambda^m = f_m(\lambda)$，则由前面的说明，此等式右边各个矩阵的和也就是（$\star$）式右边的矩阵。□

下面定理给出了矩阵特征值与矩阵范数间的一个基本关系。

定理 5 矩阵 A 的每一个特征值 λ 的模 $|\lambda|$，都不大于矩阵 A 的任何一种范数 $\|A\|$。即是说，$|\lambda| \leqslant \|A\|$。

证明 设 $\|A\| = a$。作矩阵

$$B = \frac{1}{a+\varepsilon}A \quad (\varepsilon \text{ 是任意正数}),$$

于是

$$\|B\| = \frac{1}{a+\varepsilon}\|A\| = \frac{a}{a+\varepsilon} < 1.$$

因此，当 $m \to \infty$ 时，$B^m \to 0$（矩阵）（定理 3）。但由定理 4，我们又得知矩阵 B 的所有特征值的模都小于 1，而 B 的特征值就是 $\frac{\lambda}{a+\varepsilon}$，故

$$\left|\frac{\lambda}{a+\varepsilon}\right| < 1,$$

即 $|\lambda| < a+\varepsilon$。因为正数 ε 可以任意小，因此 $|\lambda| \leqslant a = \|A\|$。从而定理 5 得证。

§4 矩阵幂级数

本节主要讨论矩阵幂级数的问题，因为在下一节我们将应用矩阵幂级数来引入矩阵函数——这与解析函数的情形相类似。在讨论幂级数之前，先简单介绍一些矩阵级数的基本概念及性质。正如前面两节那样，我们所讨论的矩阵仍然是 $n\times n$ 的复方阵。

若给定 $C^{n\times n}$ 中一方阵序列

$$A_0, A_1, A_2, \cdots, A_m, \cdots,$$

则和式

$$A_0 + A_1 + A_2 + \cdots + A_m + \cdots \tag{1}$$

称为方阵级数，也常缩写为 $\sum_{m=0}^{\infty} A_m$。令

$$S_N = \sum_{m=0}^{N} A_m,$$

若方阵序列 $\{S_N\}$ 收敛于方阵 S，则称方阵级数(1)收敛，且其和为 S，记为

$$S = \sum_{m=0}^{\infty} A_m。$$

显然 $\sum_{m=0}^{\infty} A_m$ 收敛的充要条件是 n^2 个数值级数 $\sum_{m=0}^{\infty} (A_m)_{ij}$ 收敛 $(i, j = 1, 2, \cdots, n)$。这里 $(A_m)_{ij}$ 表示 n 阶方阵 A_m 的位于第 i 行、第 j 列上元素。

又当这 n^2 个数值级数绝对收敛时，则称方阵级数(1)是绝对收敛的。

关于方阵级数的收敛问题，有下列基本性质：

1）若方阵级数 $\sum_{m=0}^{\infty} A_m$ 绝对收敛，则它一定收敛，且任意交换各项的次序所得的新级数仍收敛，和也不改变。

这由数值级数的同一性质便可推知。

2）方阵级数$\sum_{m=0}^{\infty}A_m$绝对收敛的充要条件，是对任意一种方阵范数$\|\cdot\|$，正项级数$\sum_{m=0}^{\infty}\|A_m\|$收敛。

证明 由于方阵范数的等价性，如果我们能在方阵范数$\|A\|_1=\max_{1\leqslant j\leqslant n}\sum_{i=1}^{n}|a_{ij}|$下证明这一性质，则对其他任何方阵范数的情形，也容易得知这性质也能成立。

现假设正项级数$\sum_{m=0}^{\infty}\|A_m\|_1$收敛，则由等式

$$\|A_m\|_1=\max_{1\leqslant j\leqslant n}\sum_{i=1}^{n}|(A_m)_{ij}|$$

推知

$$\|(A_m)_{ij}\|\leqslant\|A_m\|_1\quad(i,j=1,2,\cdots,n).$$

故由正项级数的比较判别法，得知级数

$$\sum_{m=0}^{\infty}|(A_m)_{ij}|\quad(i,j=1,2,3,\cdots,n)$$

都收敛，从而方阵级数

$$\sum_{m=0}^{\infty}A_m$$

绝对收敛。

反过来，若方阵级数

$$\sum_{m=0}^{\infty}A_m$$

绝对收敛，则正项级数

$$\sum_{m=0}^{\infty}|(A_m)_{ij}|\quad(i,j=1,2,\cdots,n)$$

都收敛，从而由这n^2个级数相加所构成的级数收敛，因此级数

$$\sum_{m=0}^{\infty}\left(\sum_{i=1}^{n}\sum_{j=1}^{n}|(A_m)_{ij}|\right)$$

也收敛，但是我们又有

$$\|A_m\|_1 = \max_{1\leqslant j\leqslant n}\sum_{i=1}^{n}|(A_m)_{ij}|$$

$$\leqslant\sum_{i=1}^{n}\sum_{j=1}^{n}|(A_m)_{ij}|,$$

故由正项级数的比较判别法得知级数

$$\sum_{m=0}^{\infty}\|A_m\|_1$$

收敛。证毕。

3) 若 $P,Q\in C^{n\times n}$ 为给定矩阵，如果方阵级数 $\sum\limits_{m=0}^{\infty}A_m$ 收敛（或绝对收敛），则级数 $\sum\limits_{m=0}^{\infty}PA_mQ$ 也收敛（或绝对收敛），且有等式

$$\sum_{m=0}^{\infty}PA_mQ = P(\sum_{m=0}^{\infty}A_m)Q. \tag{2}$$

证明 设 $\sum\limits_{m=0}^{\infty}A_m$ 收敛于方阵 S，即

$$\lim_{N\to\infty}\sum_{m=0}^{N}A_m = S = \sum_{m=0}^{\infty}A_m$$

而由等式 $\sum\limits_{m=0}^{N}PA_mQ=P(\sum\limits_{m=0}^{N}A_m)Q$ 取极限即得

$$\lim_{N\to 0}\sum_{m=0}^{N}PA_mQ = P(\lim_{N\to\infty}\sum_{m=0}^{N}A_m)Q = PSQ,$$

即 $\sum\limits_{m=0}^{\infty}PA_mQ$ 收敛，且有(2)式。

现设 $\sum\limits_{m=0}^{\infty}A_m$ 绝对收敛，则由性质(2)，$\sum\|A_m\|$ 也收敛。又由

$$\|PA_mQ\| \leqslant \|P\|\cdot\|A_m\|\cdot\|Q\| = \|P\|\cdot\|Q\|\cdot\|A_m\|,$$

及比较判别法，即知级数 $\sum\limits_{m=0}^{\infty}\|PA_mQ\|$ 收敛，再由性质 2)，便知方

阵级数$\sum_{m=0}^{\infty}PA_mQ$绝对收敛。证毕。

（注：基本性质3）对于P,Q分别为n维行向量与列向量的情形，可以证明也能成立。）

现在我们转到方阵幂级数的问题上来。若已给n阶复数方阵序列$\{A^m\}$及复数序列$\{C_m\}$，则方阵级数

$$\sum_{m=0}^{\infty}C_mA^m$$

称为方阵A的幂级数。

如果$\lambda_1,\lambda_2,\cdots,\lambda_n$为方阵$A\in C^{n\times n}$的全部特征值，则

$$\rho(A)=\max_{1\leqslant i\leqslant n}\{|\lambda_i|\}$$

称为A的谱半径。

定理1 若$A\in C^{n\times n}$，则对于任给正数ε，都有某一方阵范数$\|\cdot\|$，使得

$$\|A\|\leqslant\rho(A)+\varepsilon.$$

证明 对于$A\in C^{n\times n}$，必有可逆矩阵$P\in C^{n\times n}$，使A与其约当标准形J相似：

$$J=P^{-1}AP=\begin{pmatrix}t_{11} & & & \\ t_{21} & t_{22} & & \\ & \ddots & \ddots & \\ & & t_{n,n-1} & t_{nn}\end{pmatrix}$$

这里$t_{11},t_{22},\cdots,t_{nn}$是$A$的特征值，而$t_{21},\cdots,t_{n,n-1}$等于1或0，对给定的$\varepsilon>0$，取对角形矩阵

$$D=\begin{pmatrix}1 & & & & \\ & \varepsilon & & & \\ & & \varepsilon^2 & & \\ & & & \ddots & \\ & & & & \varepsilon^{n-1}\end{pmatrix}$$

显然 D 可逆，且由计算可得

$$D^{-1}JD = \begin{pmatrix} t_{11} & & & \\ \varepsilon t_{21} & t_{22} & & \\ & \ddots & \ddots & \\ & & \varepsilon t_{n,n-1} & t_{nn} \end{pmatrix}$$

对所给方阵 A，令

$$\| A \| = \| D^{-1}JD \|_{\infty},$$

可验证 $\| A \|$ 是方阵范数。回顾方阵范数 $\| B \|_{\infty}$的定义：

$$\| B \|_{\infty} = \max_{1\leqslant i\leqslant n} \sum_{j=1}^{n} |b_{ij}|,$$

从而可得

$$\| A \| = \| D^{-1}JD \|_{\infty} \leqslant \max_{1\leqslant i\leqslant n} |t_{ii}| + \max_{1\leqslant i\leqslant n-1} |t_{i+1,i}|\varepsilon$$

$$\leqslant \rho(A) + \varepsilon.$$

现在我们可以来证明本节的主要结论了：

定理 2 若复变数幂级数 $\sum_{m=0}^{\infty} C_m z^m$ 的收敛半径为 R，而方阵 $A \in C^{n\times n}$的谱半径为 $\rho(A)$，则：

(1) 当 $\rho(A)<R$ 时，方阵幂级数 $\sum_{m=0}^{\infty} C_m A^m$ 绝对收敛；

(2) 当 $\rho(A)>R$ 时，方阵幂级数 $\sum_{m=0}^{\infty} C_m A^m$ 发散。

证明 (1) 因 $\rho(A)<R$，所以总可找到正数 ε，使得 $\rho(A)+\varepsilon<R$ 仍成立。其实只须取 $0<\varepsilon<R-\rho(A)$即可。又因为幂级数 $\sum_{m=0}^{\infty} C_m z^m$ 在收敛圆 $|z|<R$ 内绝对收敛，所以正项级数

$$\sum_{m=0}^{\infty} |C_m| (\rho(A) + \varepsilon)^m$$

收敛，从而其部分和

$$S_N = \sum_{m=0}^{N} |C_m| (\rho(A) + \varepsilon)^m$$

有上界：$S_N < M$。

由定理 1，有某一方阵范数 $\|\cdot\|$，使得 $\|A\| \leqslant \rho(A) + \varepsilon$。因而

$$\begin{aligned}
\sum_{m=0}^{N} \|C_m A^m\| &\leqslant \sum_{m=0}^{N} |C_m| \cdot \|A\|^m \\
&\leqslant \sum_{m=0}^{N} |C_m| (\rho(A) + \varepsilon)^m \\
&= S_N < M.
\end{aligned}$$

所以，正项级数

$$\sum_{m=0}^{\infty} \|C_m A^m\|$$

收敛。再由收敛方阵级数的基本性质 2)，我们就得知方阵幂级数 $\sum_{m=0}^{\infty} C_m A^m$ 绝对收敛。

(2) 若 $\rho(A) > R$，设 $Ax = \lambda_j x$，且总可以取 x 为单位向量，及 $|\lambda_j| = \rho(A)$。下面用反证法证明 $\sum_{m=0}^{\infty} C_m A^m$ 发散。

如果它收敛，则由上述基本性质 3)知级数

$$\begin{aligned}
x^H \left(\sum_{m=0}^{\infty} C_m A^m\right) x &= \sum_{m=0}^{\infty} C_m x^H A^m x \\
&= \sum_{m=0}^{\infty} C_m x^H \lambda_j^m x \\
&= \sum_{m=0}^{\infty} C_m \lambda_j^m x^H x \\
&= \sum_{m=0}^{\infty} C_m \lambda_j^m \quad (\text{因 } x^H x = |x|^2 = 1)
\end{aligned}$$

也收敛。但复变数幂级数 $\sum_{m=0}^{\infty} C_m z^m$ 在收敛圆外（即当 $|z|>R$）发散。现在 $|\lambda_j|=\rho(A)>R$，故 $\sum_{m=0}^{\infty} C_m \lambda_j^m$ 又应该是发散的。这是个矛盾，因此证明了(2)。

推论 1 若复数幂级数

$$\sum_{m=0}^{\infty} C_m (z-\lambda_0)^m$$

的收敛半径是 R，则对于方阵 $A \in C^{n \times n}$，当其特征值 $\lambda_1, \lambda_2, \cdots, \lambda_n$ 满足

$$|\lambda_i - \lambda_0| < R \quad (i=1,2,\cdots,n)$$

时，方阵幂级数

$$\sum_{m=0}^{\infty} C_m (A-\lambda_0 E)^m \tag{1}$$

绝对收敛；若有某一 λ_i 使得 $|\lambda_i-\lambda_0|>R$，则方阵幂级数(1)发散。

证明 令 $B=A-\lambda_0 E$，则因 B, A 的特征值的关系为 $\lambda_B=\lambda_A-\lambda_0$，所以

$$\rho(B)=\max_{1\leqslant i\leqslant n}\{|\lambda_i-\lambda_0|\}$$

又

$$\sum_{m=0}^{\infty} C_m (A-\lambda_0 E)^m = \sum_{m=0}^{\infty} C_m B^m.$$

故由定理 2，当 $\rho(B)<R$ 时，方阵幂级数(1)绝对收敛，而条件 $|\lambda_i-\lambda_0|<R$ $(i=1,2,\cdots,n)$，当然保证 $\rho(B)<R$ 成立；当有某个 λ_i，使得 $|\lambda_i-\lambda_0|>R$ 时，则 $\rho(B)>R$，故由定理 2，此时方阵幂级数(1)发散。

推论 2 若复变数幂级数 $\sum_{m=0}^{\infty} C_m z^m$ 在整个复平面上都收敛，则对任意的方阵 $A \in C^{n \times n}$，方阵幂级数 $\sum_{m=0}^{\infty} C_m A^m$ 也收敛。

证明 取一个适当大的正数 R,使得 A 的所有特征值 $\lambda_i(i=1,2,\cdots,n)$的模均满足$|\lambda_i|<R$,由于复变数幂级数

$$\sum_{m=0}^{\infty} C_m z^m$$

在整个复平面上收敛。特别地,它在$|z|<R$ 内收敛。故由推论 1,方阵幂级数$\sum_{m=0}^{\infty} C_m A^m$ 收敛。

§5 矩阵函数

现在,我们可以像复变函数论那样,利用方阵幂级数来定义方阵函数了。

在复变函数论中,我们知道,复变数幂级数

$$\sum_{m=0}^{\infty} \frac{z^m}{m!}, \quad \sum_{m=1}^{\infty}(-1)^{m-1} \frac{z^{2m-1}}{(2m-1)!}, \quad 1+\sum_{m=1}^{\infty}(-1)^m \frac{z^{2m}}{(2m)!}$$

都在整个复平面上收敛,因而它们都有确定的和,并依次用 e^z, $\sin z$, $\cos z$ 来表示。换言之,复变数初等函数 e^z, $\sin z$ $\cos z$可用幂级数来定义:

$$e^z = \sum_{m=0}^{\infty} \frac{z^m}{m!},$$

$$\sin z = \sum_{m=1}^{\infty}(-1)^{m-1} \frac{z^{2m-1}}{(2m-1)!},$$

$$\cos z = 1+\sum_{m=1}^{\infty}(-1)^m \frac{z^{2m}}{(2m)!}.$$

因此,利用上节定理 2 的推论 2,可知对任何方阵 $A\in C^{n\times n}$,下列各方阵幂级数

$$\sum_{m=0}^{\infty} \frac{A^m}{m!}, \quad \sum_{m=1}^{\infty}(-1)^{m-1} \frac{A^{2m-1}}{(2m-1)!}, \quad E+\sum_{m=1}^{\infty}(-1)^m \frac{A^{2m}}{(2m)!}$$

都收敛。因此,我们可用它们来定义下列三个方阵函数:

$$e^A = \sum_{m=0}^{\infty} \frac{A^m}{m!},$$

$$\sin A = \sum_{m=1}^{\infty} (-1)^{m-1} \frac{A^{2m-1}}{(2m-1)!},$$

$$\cos A = E + \sum_{m=1}^{\infty} (-1)^m \frac{A^{2m}}{(2m)!},$$

分别称之为方阵 A 的指数函数、正弦函数及余弦函数．同样地，由

$$\ln(1+z) = \sum_{m=1}^{\infty} \frac{(-1)^{m-1}}{m} z^m \quad (|z| < 1),$$

$$(1+z)^a = 1 + \sum_{m=1}^{\infty} \frac{a(a-1)\cdots(a-m+1)}{m!} z^m \quad (|z| < 1)$$

(a 为任意实数)

可定义方阵函数 $\ln(E+A)$ 及 $(E+A)^a$ 为：

$$\ln(E+A) = \sum_{m=1}^{\infty} \frac{(-1)^{m-1}}{m} A^m \quad (\rho(A) < 1),$$

$$(E+A)^a = E + \sum_{m=1}^{\infty} \frac{a(a-1)\cdots(a-m+1)}{m!} A^m \quad (\rho(A) < 1).$$

(特别地，$(E-A)^{-1} = \sum_{m=0}^{\infty} A^m$ 是个有重要应用的例子.)一般地，若复变数幂级数

$$\sum_{m=0}^{\infty} C_m z^m$$

的收敛半径为 R，其和为 $f(z)$，即

$$f(z) = \sum_{m=0}^{\infty} C_m z^m \quad (|z| < R).$$

则由上节定理 2，就可定义矩阵函数

$$f(A) = \sum_{m=0}^{\infty} C_m A^m \quad (\rho(A) < R; A \in C^{n\times n}).$$

现在的问题是如何求出矩阵函数 $f(A)$ 来．本节主要讨论用矩阵 A

的标准形来计算矩阵函数 $f(A)$ 的方法，同时也简单介绍一下用多项式来表示矩阵函数 $f(A)$ 的方法，首先我们来证明两个有关的定理。

定理 1 若对任一方阵 X，幂级数 $\sum_{m=0}^{\infty} C_m X^m$ 都收敛，其和为

$$f(X) = \sum_{m=0}^{\infty} C_m X^m,$$

则当 X 为分块对角形矩阵

$$X = \begin{bmatrix} X_1 & & & \\ & X_2 & & \\ & & \ddots & \\ & & & X_k \end{bmatrix}$$

时，即有

$$f(X) = \begin{bmatrix} f(X_1) & & & \\ & f(X_2) & & \\ & & \ddots & \\ & & & f(X_k) \end{bmatrix}$$

证明 因为

$$f(X) = \lim_{N\to\infty} \sum_{m=0}^{N} C_m \begin{bmatrix} X_1^m & & & \\ & X_2^m & & \\ & & \ddots & \\ & & & X_k^m \end{bmatrix}$$

$$= \begin{bmatrix} \lim\limits_{N\to\infty} \sum\limits_{m=0}^{N} C_m X_1^m & & \\ & \ddots & \\ & & \lim\limits_{N\to\infty} \sum\limits_{m=0}^{N} C_m X_k^m \end{bmatrix}$$

$$=\begin{bmatrix} f(X_1) & & & \\ & f(X_2) & & \\ & & \ddots & \\ & & & f(X_k) \end{bmatrix}$$

定理 1 得证。

定理 2　若

$$f(z)=\sum_{m=0}^{\infty}C_m z^m \quad (|z|<R)$$

是收敛半径为 R 的复变数幂级数，又

$$J_0=\begin{bmatrix} \lambda_0 & & & \\ 1 & \lambda_0 & & \\ & \ddots & \ddots & \\ & & 1 & \lambda_0 \end{bmatrix}$$

是 n 阶约当块，则当 $|\lambda_0|<R$ 时，方阵幂级数

$$\sum_{m=0}^{\infty}C_m J_0^m$$

绝对收敛，且其和为

$$f(J_0)=\begin{bmatrix} f(\lambda_0) & & & & \\ f'(\lambda_0) & \ddots & & & \\ \frac{1}{2!}f''(\lambda_0) & \ddots & \ddots & & \\ \vdots & \ddots & \ddots & \ddots & \\ \frac{1}{(n-1)!}f^{(n-1)}(\lambda_0) & \cdots & \frac{1}{2!}f''(\lambda_0) & f'(\lambda_0) & f(\lambda_0) \end{bmatrix} \quad (1)$$

证明　由于 J_0 的特征值为 λ_0，且为 n 重根，故 $\rho(J_0)=|\lambda_0|$，由上节的定理 2，便得知当 $|\lambda_0|<R$ 时，方阵幂级数 $\sum_{m=0}^{\infty}C_m J_0^m$ 绝对收敛。因此，只须证明这级数的和 $f(J_0)=\sum_{m=0}^{\infty}C_m J_0^m$ 恰为(1)式右边的

矩阵就行了。首先，由定义我们有

$$f(J_0)=\lim_{N\to\infty}f_N(J_0)=\lim_{N\to\infty}\sum_{m=0}^{N}C_mJ_0^m,$$

应用上节定理 4 后面的注，并经简化可得

$$f_N(J_0)=\begin{bmatrix} f_N(\lambda_0) & & & & \\ f'_N(\lambda_0) & \ddots & & & \\ \frac{1}{2!}f''_N(\lambda_0) & \ddots & \ddots & & \\ \vdots & \ddots & \ddots & \ddots & \\ \frac{1}{(n-1)!}f_N^{(n-1)}(\lambda_0) & \cdots & \frac{1}{2!}f''_N(\lambda_0) & f'_N(\lambda_0) & f_N(\lambda_0) \end{bmatrix} \quad (2)$$

在这里，

$$\begin{cases} f_N(\lambda_0)=\sum_{m=0}^{N}C_m\lambda_0^m \\ f'_N(\lambda_0)=\sum_{m=1}^{N}mC_m\lambda_0^{m-1} \\ f''_N(\lambda_0)=\sum_{m=2}^{N}m(m-1)C_m\lambda_0^{m-2} \\ \cdots\cdots\cdots\cdots\cdots\cdots \\ f_N^{(n-1)}(\lambda_0)=\sum_{m=n-1}^{N}m(m-1)\cdots(m-n+2)C_m\lambda_0^{m-n+1} \end{cases} \quad (3)$$

由于当 $|z|<R$ 时，级数 $\sum_{m=0}^{\infty}C_mz^m$ 绝对收敛，且它的和为 $f(z)$，由复变函数论知道，$f(z)$ 是 $|z|<R$ 内的解析函数，它有任意阶的导数：

$$f^{(k)}(z)=\sum_{m=k}^{\infty}m(m-1)\cdots(m-k+1)C_mz^{m-k}\quad(|z|<R),$$

$$(k=1,2,\cdots)。$$

由于 $|\lambda_0|<R$，故当 $z=\lambda_0$ 时，这 k 阶导数等式也成立。从而(3)式中各个等式当 $N\to\infty$ 时的极限都存在，且

$$
\begin{cases}
\lim\limits_{N\to\infty} f_N(\lambda_0)=\lim\limits_{N\to\infty}\sum\limits_{m=0}^{N} C_m\lambda_0^m=f(\lambda_0)\\
\lim\limits_{N\to\infty} f'_N(\lambda_0)=\lim\limits_{N\to\infty}\sum\limits_{m=1}^{N} mC_m\lambda_0^{m-1}=f'(\lambda_0)\\
\lim\limits_{N\to\infty} f''_N(\lambda_0)=\lim\limits_{N\to\infty}\sum\limits_{m=2}^{N} m(m-1)C_m\lambda_0^{m-2}=f''(\lambda_0)\\
\lim\limits_{N\to\infty} f_N^{(n-1)}(\lambda_0)=\lim\limits_{N\to\infty}\sum\limits_{m=n-1}^{N} m(m-1)\cdots(m-n+2)C_m\lambda^{m-n+1}=f^{(n-1)}(\lambda_0)
\end{cases}
\tag{4}
$$

令 $N\to\infty$，对(2)式两边取极限(并应用(4)中各式)，便得所证。

现在，我们可以用矩阵 A 的标准形来计算矩阵函数 $f(A)$ 了。我们分两个情形来讨论。

1）若 A 相似于对角形矩阵：

$$
A=P\begin{bmatrix}\lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n\end{bmatrix}P^{-1}
$$

简记为 $A=PJP^{-1}$。这里 λ_i 是 A 的特征值($i=1,2,\cdots,n$)。

由上节定理 2，如果复变数幂级数

$$
\sum_{m=0}^{\infty} C_m z^m
$$

的收敛半径为 R，则当 $|z|<R$ 时，此级数绝对收敛，其和设为 $f(z)$。又当方阵 A 的谱半径 $\rho(A)<R$ 时，方阵幂级数

$$
\sum_{m=0}^{\infty} C_m A^m
$$

也绝对收敛，且其和为

$$
f(A)=\sum_{m=0}^{\infty} C_m A^m=f(PJP^{-1})
$$

$$= \sum_{m=0}^{\infty} C_m (PJP^{-1})^m$$

$$= \sum_{m=0}^{\infty} C_m PJ^m P^{-1} = P\left(\sum_{m=0}^{\infty} C_m J^m\right) P^{-1}$$

$$= Pf(J)P^{-1}.$$

由定理1得

$$f(A) = P \begin{bmatrix} f(\lambda_1) & & & \\ & f(\lambda_2) & & \\ & & \ddots & \\ & & & f(\lambda_n) \end{bmatrix} P^{-1} \quad (*)$$

并且可见 $f(A)$ 的特征值为 $f(\lambda_1), f(\lambda_2), \cdots, f(\lambda_n)$。因而，在 A 相似于对角形矩阵时，与复变数幂级数

$$f(z) = \sum_{m=0}^{\infty} C_m z^m \quad (|z| < R)$$

相应的矩阵函数

$$f(A) = \sum_{m=0}^{\infty} C_m A^m \quad (\rho(A) < R)$$

也相似于一个对角形矩阵，而矩阵函数 $f(A)$ 就是一个由（*）式确定的矩阵。

特别地，我们有

$$e^A = P \begin{bmatrix} e^{\lambda_1} & & & \\ & e^{\lambda_2} & & \\ & & \ddots & \\ & & & e^{\lambda_n} \end{bmatrix} P^{-1}$$

$$\sin A = P \begin{bmatrix} \sin\lambda_1 & & & \\ & \sin\lambda_2 & & \\ & & \ddots & \\ & & & \sin\lambda_n \end{bmatrix} P^{-1}$$

$$\cos A = P\begin{bmatrix} \cos\lambda_1 & & & \\ & \cos\lambda_2 & & \\ & & \ddots & \\ & & & \cos\lambda_n \end{bmatrix}P^{-1}$$

这只须分别取 $f(z)=e^z, \sin z, \cos z$，并假定 A 是相似于对角形矩阵的（此外，对 A 再无其他要求）。由此可以看出 $e^A, \sin A, \cos A$ 这三个矩阵的特征值是什么。

例1 设

$$A = \begin{pmatrix} 0 & 1 \\ 0 & 2 \end{pmatrix}$$

求 $e^A, \sin A, \cos A$。

解 因

$$|\lambda E - A| = \begin{vmatrix} \lambda & -1 \\ 0 & \lambda - 2 \end{vmatrix} = \lambda \cdot (\lambda - 2)$$

故 A 有不同的特征值 $\lambda_1=0, \lambda_2=2$。所以 A 能相似于对角形，不难求得相应于两个特征值的特征向量分别为：

$$\alpha_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \qquad \alpha_2 = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$$

令

$$P = \begin{pmatrix} 1 & 1 \\ 0 & 2 \end{pmatrix}$$

则有

$$P^{-1} = \frac{1}{2}\begin{pmatrix} 2 & -1 \\ 0 & 1 \end{pmatrix} \qquad A = P\begin{pmatrix} 0 & \\ & 2 \end{pmatrix}P^{-1}.$$

$$e^A = P\begin{pmatrix} 1 & \\ & e^2 \end{pmatrix}P^{-1} = \begin{bmatrix} 1 & -\frac{1}{2} + \frac{1}{2}e^2 \\ 0 & e^2 \end{bmatrix}$$

$$\sin A = \begin{pmatrix}1 & 1\\ 0 & 2\end{pmatrix}\begin{pmatrix}0 & \\ & \sin 2\end{pmatrix}\begin{bmatrix}1 & -\frac{1}{2}\\ 0 & \frac{1}{2}\end{bmatrix}$$

$$= \begin{bmatrix}0 & \frac{1}{2}\sin 2\\ 0 & \sin 2\end{bmatrix}$$

$$\cos A = \begin{pmatrix}1 & 1\\ 0 & 2\end{pmatrix}\begin{pmatrix}1 & \\ & \cos 2\end{pmatrix}\begin{bmatrix}1 & -\frac{1}{2}\\ 0 & \frac{1}{2}\end{bmatrix}$$

$$= \begin{bmatrix}1 & -\frac{1}{2}+\frac{1}{2}\cos 2\\ 0 & \cos 2\end{bmatrix}$$

在实用上，遇到的往往不是常数矩阵 A 的矩阵函数 $f(A)$，而是变量 t 的函数矩阵 At 的矩阵函数 $f(At)$。

注意到当

$$A = PJP^{-1} = P\begin{bmatrix}\lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n\end{bmatrix}P^{-1}$$

时，则

$$At = P(Jt)P^{-1} = P\begin{bmatrix}\lambda_1 t & & \\ & \lambda_2 t & \\ & & \lambda_n t\end{bmatrix}P^{-1}$$

因而类似上述的推导，即有

$$f(At) = P\begin{bmatrix} f(\lambda_1 t) & & & \\ & f(\lambda_2 t) & & \\ & & \ddots & \\ & & & f(\lambda_n t) \end{bmatrix} P^{-1}$$

特别地，我们有

$$e^{At} = P\begin{bmatrix} e^{\lambda_1 t} & & & \\ & e^{\lambda_2 t} & & \\ & & \ddots & \\ & & & e^{\lambda_n t} \end{bmatrix} P^{-1},$$

$$\sin At = P\begin{bmatrix} \sin\lambda_1 t & & & \\ & \sin\lambda_2 t & & \\ & & \ddots & \\ & & & \sin\lambda_n t \end{bmatrix} P^{-1}$$

$$\cos At = P\begin{bmatrix} \cos\lambda_1 t & & & \\ & \cos\lambda_2 t & & \\ & & \ddots & \\ & & & \cos\lambda_n t \end{bmatrix} P^{-1}$$

例如，对上述例 1 中的矩阵 A，则有

$$e^{At} = \begin{pmatrix} 1 & 1 \\ 0 & 2 \end{pmatrix}\begin{pmatrix} 1 & \\ & e^{2t} \end{pmatrix}\begin{bmatrix} 1 & -\frac{1}{2} \\ 0 & \frac{1}{2} \end{bmatrix}$$

$$= \begin{bmatrix} 1 & \frac{1}{2}(-1 + e^{2t}) \\ 0 & e^{2t} \end{bmatrix}$$

2）当 A 不能与对角形矩阵相似，这时，A 必可与其约当标准形相似：

$$A = P\begin{bmatrix} J_1(\lambda_1) & & & \\ & J_2(\lambda_2) & & \\ & & \ddots & \\ & & & J_k(\lambda_k) \end{bmatrix}P^{-1}$$

其中约当块

$$J_i(\lambda_i) = \begin{bmatrix} \lambda_i & & & \\ 1 & \lambda_i & & \\ & \ddots & \ddots & \\ & & 1 & \lambda_i \end{bmatrix}_{n_i \times n_i}$$

由初级因子$(\lambda-\lambda_i)^{n_i}$所决定。又 $n_1+n_2+\cdots+n_k=n$。

因此,应用定理 1 我们有:

$$f(A) = \sum_{m=0}^{\infty} C_m A^m$$

$$= \sum_{m=0}^{\infty} C_m P\begin{bmatrix} J_1^m(\lambda_1) & & & \\ & J_2^m(\lambda_2) & & \\ & & \ddots & \\ & & & J_k^m(\lambda_k) \end{bmatrix}P^{-1}$$

$$= P(\sum_{m=0}^{\infty} C_m \begin{bmatrix} J_1^m & & & \\ & J_2^m & & \\ & & \ddots & \\ & & & J_k^m \end{bmatrix})P^{-1}$$

$$= P\begin{bmatrix} f(J_1) & & & \\ & f(J_2) & & \\ & & \ddots & \\ & & & f(J_k) \end{bmatrix} p^{-1}$$

再参照定理 2 的公式计算出每个 $f(J_i)$ $(i=1,2,\cdots,k)$,从而便可得到 $f(A)$。

类似地有

$$f(At) = P\begin{bmatrix} f(J_1t) & & & \\ & f(J_2t) & & \\ & & \ddots & \\ & & & f(J_kt) \end{bmatrix}P^{-1}$$

例2 设

$$A = \begin{bmatrix} -1 & 1 & 0 \\ -4 & 3 & 0 \\ 1 & 0 & 2 \end{bmatrix}$$

求 e^A(见第三章第三节)。

解 我们已求得

$$P = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 2 \\ 1 & -1 & -1 \end{bmatrix} \qquad P^{-1}AP = J = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix}$$

且易得

$$P^{-1} = \begin{bmatrix} -1 & 1 & 1 \\ -2 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}$$

因此,(考虑 $f(z)=e^z=\sum_{m=0}^{\infty}\frac{z^m}{m!}$)

$$e^A = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 2 \\ 1 & -1 & -1 \end{bmatrix}\begin{bmatrix} e^2 & 0 & 0 \\ 0 & e^1 & 0 \\ 0 & e^1 & e^1 \end{bmatrix}\begin{bmatrix} -1 & 1 & 1 \\ -2 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}$$

$$= \begin{bmatrix} 0 & e & e \\ 0 & 3e & 2e \\ e^2 & -2e & -e \end{bmatrix}\begin{bmatrix} -1 & 1 & 1 \\ -2 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}$$

$$= \begin{bmatrix} -e & e & 0 \\ -4e & 3e & 0 \\ 3e-e^2 & e^2-2e & e^2 \end{bmatrix}$$

若在计算 e^A 的第一个等式中,用

$$\begin{bmatrix} e^{2t} & 0 & 0 \\ 0 & e^t & 0 \\ 0 & te^t & e^t \end{bmatrix}$$

去代替

$$\begin{bmatrix} e^2 & 0 & 0 \\ 0 & e & 0 \\ 0 & e & e \end{bmatrix}$$

则读者还可求得含变量 t 的矩阵函数 e^{At}。

以上讨论了用方阵幂级数来表示方阵函数的方法及具体计算问题。从所举例子可以看到，对一个比较简单的方阵 A，即使要求出最简单的矩阵函数 e^A，$\sin A$ 来也还是比较繁琐的。下面介绍的多项式表示法，比起上述方法来要简便些。但是，这一方法的理论推导则相当繁复，所以这里不作推导，我们只将结果介绍给读者。

若 $f(\lambda)$是 $l(>m)$次多项式，$\varphi(\lambda)$是方阵 A 的最小多项式，它的次数为 m。以 $\varphi(\lambda)$去除 $f(\lambda)$即得

$$f(\lambda) = \varphi(\lambda)q(\lambda) + r(\lambda),$$

这里余式 $r(\lambda)=0$ 或比 $\varphi(\lambda)$有更低的次数。因此，

$$f(A) = \varphi(A)q(A) + r(A) = r(A).$$

由此可见，次数高于 m 的任一多项式 $f(A)$（矩阵 A 的一个矩阵多项式）都可以化为次数$\leqslant m-1$ 的 A 的多项式 $r(A)$来计算。换言之，若 $f(A)$是一个 l 次的矩阵多项式

$$f(A) = a_0E + a_1A + a_2A^2 + \cdots + a_lA^l \quad (l > m),$$

而 $\varphi(\lambda)$是 A 的最小多项式，其次数为 m，则有

$$f(A) = r(A) = \alpha_0E + \alpha_1A + \alpha_2A^2 + \cdots + \alpha_{m-1}A^{m-1}.$$

即 $f(A)$可以表示为一个关于 A 的次数$\leqslant m-1$ 的矩阵多项式 $r(A)$。

把这一思想推广到由矩阵幂级数确定的矩阵函数 $f(A)$上，那

就是所谓矩阵函数的多项式表示问题。我们把它叙述为如下的定理。

定理 3 设 n 阶方阵 A 的最小多项式为 m 次多项式

$$\varphi(\lambda)=(\lambda-\lambda_1)^{n_1}(\lambda-\lambda_2)^{n_2}\cdots(\lambda-\lambda_s)^{n_s},$$

其中 $\lambda_1,\lambda_2,\cdots,\lambda_s$ 是 A 的所有互不相同的特征值。又与收敛的复变数幂级数 $f(z)=\sum_{k=0}^{\infty}C_kz^k$ 相应的

$$f(A)=\sum_{k=0}^{\infty}C_kA^k$$

是 A 的收敛幂级数，则矩阵函数 $f(A)$ 可以表示成 A 的 $m-1$ 次多项式

$$f(A)=a_0E+a_1A+a_2A^2+\cdots+a_{m-1}A^{m-1},$$

系数 $a_0,a_1,a_2,\cdots,a_{m-1}$ 由下列方程组的解给出：

$$\begin{cases}a_0+a_1\lambda_i+a_2\lambda_i^2+\cdots+a_{m-1}\lambda_i^{m-1}=f(\lambda_i),\\ a_1+2a_2\lambda_i+\cdots+(m-1)a_{m-1}\lambda_i^{m-2}=f'(\lambda_i),\\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\\ (n_i-1)!\,a_{n_i-1}+\cdots+(m-1)\cdots(m-n_i+1)a_{m-1}\lambda_i^{m-n_i}=f^{(n_i-1)}(\lambda_i).\end{cases}$$

$(\forall\ \lambda_i,i=1,2,\cdots,s)$

我们用例子说明这一方法。

例 3 用定理 3 提供的方法，对前述例 1 的矩阵 A，计算 e^A。

解 由 A 的特征多项式

$$|\lambda E-A|=\lambda(\lambda-2)$$

及最小多项式的性质，可知 A 的最小多项式 $\varphi(\lambda)=\lambda(\lambda-2)$。这是 $m=2$ 次多项式，所以设

$$e^A=a_0E+a_1A\ .$$

对 $\lambda_1=0$ 及 $\lambda_2=2$，按上法作方程组：

$$\begin{cases}a_0+a_1\cdot 0=e^0,\\ a_0+a_1\cdot 2=e^2.\end{cases}$$

由此得 $a_0=1$，$a_1=\frac{1}{2}(e^2-1)$，于是有

$$
\begin{aligned}
e^A &= E+\frac{1}{2}(e^2-1)A \\
&= \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + \frac{1}{2}(e^2-1)\begin{pmatrix} 0 & 1 \\ 0 & 2 \end{pmatrix} \\
&= \begin{bmatrix} 1 & \frac{1}{2}(e^2-1) \\ 0 & e^2 \end{bmatrix}
\end{aligned}
$$

对于含变数 t 的方阵函数 $f(At)$，有类似的方法，这时，

$$f(At)=a_0(t)E+a_1(t)A+a_2(t)A^2+\cdots+a_{m-1}(t)A^{m-1},$$

其中 $a_i(t)$ $(i=1,2,\cdots,m-1)$ 是 t 的函数，而确定这些 $a_i(t)$ 的方程组为

$$
\begin{cases}
a_0(t)+a_1(t)\lambda_i+\cdots+a_{m-1}(t)\lambda_i^{m-1}=f(\lambda_i t) \\
a_1(t)+2a_2(t)\lambda_i+\cdots+(m-1)a_{m-1}(t)\lambda_i^{m-2}=\left.\dfrac{\mathrm{d}f(\lambda t)}{\mathrm{d}\lambda}\right|_{\lambda=\lambda_i} \\
\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\
(n_i-1)!\,a_{n_i-1}(t)+\cdots+(m-1)(m-2)\cdots(m-n_i+1)a_{m-1}(t)\lambda_i^{m-n_i} \\
=\left.\dfrac{\mathrm{d}^{(n_i-1)}f(\lambda t)}{\mathrm{d}\lambda^{n_i-1}}\right|_{\lambda=\lambda_i}
\end{cases}
$$

$(\forall\ \lambda_i; i=1,2,\cdots,s)$

例 4 计算 e^{At}，$\sin At$，其中

$$A=\begin{bmatrix} 2 & 1 & 4 \\ 0 & 2 & 0 \\ 0 & 3 & 1 \end{bmatrix}$$

解 特征多项式

$$|\lambda E-A|=\begin{vmatrix} \lambda-2 & -1 & -4 \\ 0 & \lambda-2 & 0 \\ 0 & -3 & \lambda-1 \end{vmatrix}=(\lambda-1)(\lambda-2)^2$$

由于$(\lambda-1)(\lambda-2)$不是A的零化多项式,所以A的最小多项式为

$$\varphi(\lambda)=(\lambda-1)(\lambda-2)^2.$$

又记

$$f_1(At)=e^{At},\quad f_1(\lambda t)=e^{\lambda t},$$

$$f_2(At)=\sin At,\quad f_2(\lambda t)=\sin\lambda t\qquad \lambda_1=2,\lambda_2=1.$$

因$\varphi(\lambda)$为3次多项式,故设

$$e^{At}=a_0(t)E+a_1(t)A+a_2(t)A^2,$$

由此得方程组

$$\begin{cases}a_0(t)+a_1(t)\lambda_1+a_2(t)\lambda_1^2=f_1(\lambda_1 t)\\ a_1(t)+2a_2(t)\lambda_1=\dfrac{d}{d\lambda}f_1(\lambda t)\Big|_{\lambda=\lambda_1}\\ a_0(t)+a_1(t)\lambda_2+a_2(t)\lambda_2^2=f_1(\lambda_2 t)\end{cases}$$

即

$$\begin{cases}a_0(t)+2a_1(t)+4a_2(t)=e^{2t}\\ a_1(t)+4a_2(t)=te^{2t}\\ a_0(t)+a_1(t)+a_2(t)=e^t\end{cases}$$

由此求得

$$a_0(t)=4e^t-3e^{2t}+2te^{2t}$$

$$a_1(t)=-4e^t+4e^{2t}-3te^{2t}$$

$$a_2(t)=e^t-e^{2t}+te^{2t}$$

将其代入所设式子中,可以算出

$$e^{At}=\begin{bmatrix}e^{2t} & 12e^t-12e^{2t}+13te^{2t} & -4e^t+4e^{2t}\\ 0 & e^{2t} & 0\\ 0 & -3e^t+3e^{2t} & e^t\end{bmatrix}$$

类似地,设

$$\sin At=a_0(t)E+a_1(t)A+a_2(t)A^2\quad(\because m-1=2)$$

并由此作方程组

$$\begin{cases} a_0(t) + a_1(t)\lambda_1 + a_2(t)\lambda_1^2 = f_2(\lambda_1 t) \\ a_1(t) + 2a_2(t)\lambda_1 = \dfrac{d}{d\lambda} f_2(\lambda t)\Big|_{\lambda=\lambda_1} \\ a_0(t) + a_1(t)\lambda_2 + a_2(t)\lambda_2^2 = f_2(\lambda_2 t) \end{cases}$$

即

$$\begin{cases} a_0(t) + 2a_1(t) + 4a_2(t) = \sin 2t \\ a_1(t) + 4a_2(t) = t\cos 2t \\ a_0(t) + a_1(t) + a_2(t) = \sin t \end{cases}$$

由此解出

$$a_0(t) = 4\sin t - 3\sin 2t + 2t\cos 2t,$$

$$a_1(t) = -4\sin t + 4\sin 2t - 3t\cos 2t,$$

$$a_2(t) = \sin t - \sin 2t + t\cos 2t.$$

于是求得

$$\sin At = \begin{bmatrix} \sin 2t & 12\sin t - 12\sin 2t + 13t\cos 2t & -4\sin t + 4\sin 2t \\ 0 & \sin 2t & 0 \\ 0 & -3\sin t + 3\sin 2t & \sin t \end{bmatrix}$$

§6 矩阵的微分与积分

在应用中,矩阵函数与函数矩阵的微、积分常常是同时出现的,因此我们学习了矩阵函数的计算以后,还需学习函数矩阵的微分、积分。比起前者来,后者要简单、容易得多了,它只是一般微积分概念法则的一种形式的推广。

若 $A(z)=(a_{ij}(z))_{m\times n}$ 的每个元素 $a_{ij}(z)$ 都是复变量 z 的函数,且都在 $z=z_0$ 或变量 z 的某个区域内可导,则定义 $A(z)$ 的导数为:

$$\frac{d}{dz}A(z) = \left(\frac{d}{dz}a_{ij}(z)\right)_{m\times n}, \text{或 } A'(z) = (a'_{ij}(z))_{m\times n}.$$

换句话说,矩阵 $A(z)$ 的导数定义为由它的每个元素的导数作成的矩阵。例如

$$A(z)=\begin{pmatrix} z^2+1 & \sin z \\ 3 & e^z \end{pmatrix} \qquad \frac{d}{dz}A(z)=\begin{pmatrix} 2z & \cos z \\ 0 & e^z \end{pmatrix}$$

性质：

1) $[A(z)+B(z)]'=A'(z)+B'(z)$；

2) $[A(z)\cdot B(z)]'=A'(z)\cdot B(z)+A(z)\cdot B'(z)$。

特别，当 C 为常数矩阵时，有 $C'=0$（矩阵），且

$$[C\cdot A(z)]'=C\cdot A'(z)。$$

3) 如 $A(u)=(a_{ij}(u))_{m\times n}$ 及变量 z 的函数 $u=f(z)$ 都可导，则

$$\frac{d}{dz}A(u)=\frac{dA(u)}{du}\cdot\frac{du}{dz}.$$

以上简单性质，由定义都不难证明，兹不赘述。

4) 若 n 阶函数矩阵 $A(z)$ 可逆，且 $A(z)$ 及其逆矩阵 $A^{-1}(z)$ 都可导，则

$$\frac{d}{dz}A^{-1}(z)=-A^{-1}(z)\cdot\frac{d}{dz}A(z)\cdot A^{-1}(z).$$

证明 因

$$A^{-1}(z)\cdot A(z)=E,$$

两边对 z 求导，即有

$$\frac{d}{dz}[A^{-1}(z)\cdot A(z)]=\frac{dA^{-1}(z)}{dz}\cdot A(z)+A^{-1}(z)\cdot\frac{dA(z)}{dz}$$
$$=0（矩阵）。$$

所以

$$\frac{dA^{-1}(z)}{dz}=-A^{-1}(z)\cdot\frac{dA(z)}{dz}\cdot A^{-1}(z).$$

类似地，可定义函数矩阵对数值变量的积分。

若函数矩阵 $A(x)=(a_{ij}(x))_{m\times n}$ 的每个元素都是实变量 x 的函数，且都在 $[a,b]$ 上可积。则 $A(x)$ 的定积分与不定积分定义如下：

$$\int_a^b A(x)\mathrm{d}x = \begin{bmatrix} \int_a^b a_{11}(x)\mathrm{d}x & \cdots & \int_a^b a_{1n}(x)\mathrm{d}x \\ \cdots & \cdots & \cdots \\ \int_a^b a_{m1}(x)\mathrm{d}x & \cdots & \int_a^b a_{mn}(x)\mathrm{d}x \end{bmatrix}$$

$$\int A(x)\mathrm{d}x = \begin{bmatrix} \int a_{11}(x)\mathrm{d}x & \cdots & \int a_{1n}(x)\mathrm{d}x \\ \cdots & \cdots & \cdots \\ \int a_{m1}(x)\mathrm{d}x & \cdots & \int a_{mn}(x)\mathrm{d}x \end{bmatrix}$$

简言之，矩阵的积分定义为每个元素的积分。

易证以下各个性质成立：

1) $\int A^T(x)\mathrm{d}x = (\int A(x)\mathrm{d}x)^T$；

2) $\int [aA(x) + bB(x)]\mathrm{d}x = a\int A(x)\mathrm{d}x + b\int B(x)\mathrm{d}x$

(a,b 为非零实数)；

3) $\int C\cdot A(x)\mathrm{d}x = C\int A(x)\mathrm{d}x$ （C 为非零常数矩阵）；

4) $\int A(x)\cdot B'(x)\mathrm{d}x = A(x)B(x) - \int A'(x)\cdot B(x)\mathrm{d}x$。

对上述的矩阵微分、积分概念，还可以作些推广，比如可定义 $A(x)$ 的广义积分与拉氏变换等。但正如以上所讲的情况那样，这主要还是一种约定，实质性的东西不多，而在应用中遇到的时候，按所给出的定义去理解就行了，故在此不多讨论。

§7 常用矩阵函数的性质

在矩阵函数的应用中，常要涉及到矩阵函数的一些简单性质。这些性质，一方面从形式上看与常见函数性质相类似，且也是一种

很自然的推广；另一方面，这些性质也有一些值得注意的地方，用时都应谨慎从事。

以下是常用矩阵函数的一些基本性质，所讨论的矩阵 A,B 都是 n 阶复数方阵。

1) $\frac{\mathrm{d}}{\mathrm{d}t}e^{At}=Ae^{At}=e^{At}A$。

证明 因为

$$e^{At}=\sum_{m=0}^{\infty}\frac{1}{m!}(At)^m=\sum_{m=0}^{\infty}\frac{1}{m!}t^mA^m,$$

故

$$(e^{At})_{ij}=\sum_{m=0}^{\infty}\frac{1}{m!}(t^mA^m)_{ij}$$

对任何 t 值都收敛，因而可以逐项微分，于是有

$$\frac{\mathrm{d}}{\mathrm{d}t}(e^{At})_{ij}=\sum_{m=1}^{\infty}\frac{1}{(m-1)!}t^{m-1}(A^m)_{ij},$$

从而

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}e^{At}&=\sum_{m=1}^{\infty}\frac{1}{(m-1)!}t^{m-1}A^m\\&=A\left(\sum_{m=1}^{\infty}\frac{1}{(m-1)!}(At)^{m-1}\right)\\&=A\left(\sum_{k=0}^{\infty}\frac{1}{k!}(At)^k\right)=Ae^{At}.\end{aligned}$$

同理可得

$$\frac{\mathrm{d}}{\mathrm{d}t}e^{At}=e^{At}A.$$

由此可见，A 与 e^{At}是可交换的，它还可以推广为：

2) 若 $AB=BA$，则

$$e^{At}B=Be^{At}.$$

证明 由 $AB=BA$，首先易得 $A^mB=BA^m$，因此有

$$e^{At}B=(\sum_{m=0}^{\infty}\frac{1}{m!}A^m t^m)B$$

$$=\sum_{m=0}^{\infty}\frac{1}{m!}t^m A^m B$$

$$=\sum_{m=0}^{\infty}\frac{t^m}{m!}BA^m$$

$$=B(\sum_{m=0}^{\infty}\frac{1}{m!}(At)^m)$$

$$=Be^{At}。$$

3）若 $AB=BA$，则

$$e^A\cdot e^B=e^B\cdot e^A=e^{A+B}。$$

证明 令

$$C(t)=e^{(A+B)t}\cdot e^{-At}\cdot e^{-Bt},$$

则有

$$\frac{d}{dt}C(t)=(A+B)e^{(A+B)t}\cdot e^{-At}\cdot e^{-Bt}$$

$$-e^{(A+B)t}\cdot Ae^{-At}\cdot e^{-Bt}-e^{(A+B)t}\cdot e^{-At}\cdot Be^{-Bt}$$

$$=0(矩阵)。$$

故 C（t）为常数矩阵，因而

$$C(t)=C(1)=C(0)=e^0\cdot e^0\cdot e^0=(e^0)^3,$$

（这里的 0 是零矩阵，下同。）

但是

$$e^0=P\begin{bmatrix}e^0 & & & \\ & e^0 & & \\ & & \ddots & \\ & & & e^0\end{bmatrix}P^{-1}=PEP^{-1}=E$$

所以 C（1）$=C$（0）$=E$，故

$$e^{A+B}\cdot e^{-A}\cdot e^{-B}=E. \qquad (*)$$

特别地，取 $B=-A$，则有

$$e^0 \cdot e^{-A} \cdot e^A = E.$$

由于 $e^0=E$，因此得 $e^{-A} \cdot e^A=E$，即有

$$(e^A)^{-1} = e^{-A}。$$

这对任意方阵 A 均成立，因此，亦有 $(e^B)^{-1}=e^{-B}$。再由（*）式即有

$$e^{A+B} \cdot (e^A)^{-1} \cdot (e^B)^{-1} = E.$$

两边右乘 $e^B \cdot e^A$ 即得

$$e^{A+B} = e^B \cdot e^A,$$

但 $e^{A+B}=e^{B+A}=e^A \cdot e^B$（由上式），于是得：

$$e^A \cdot e^B = e^B \cdot e^A = e^{A+B}$$

利用绝对收敛级数的性质，容易推得

4）$e^{iA}=\cos A+i\sin A$,

$\cos A=\frac{1}{2}(e^{iA}+e^{-iA})$,

$\sin A=\frac{1}{2i}(e^{iA}-e^{-iA})$,

$\cos(-A)=\cos A$,

$\sin(-A)=-\sin A$.

5）当 $AB=BA$ 时，则有

$$\cos(A+B) = \cos A\cos B - \sin A\sin B,$$

$$\sin(A+B) = \sin A\cos B + \cos A\sin B.$$

证明 由性质 3）与 4），则有

$$\begin{aligned}
&\sin A\cos B + \cos A\sin B \\
&= \frac{1}{2i}(e^{iA} - e^{-iA}) \cdot \frac{1}{2}(e^{iB} + e^{-iB}) \\
&\quad + \frac{1}{2}(e^{iA} + e^{-iA}) \cdot \frac{1}{2i}(e^{iB} - e^{-iB}) \\
&= \frac{1}{2i}[e^{i(A+B)} - e^{-i(A+B)}]
\end{aligned}$$

$$=\sin(A+B).$$

同理可证另一等式．当 $A=B$ 时，可得

$$\cos 2A=\cos^2 A-\sin^2 A,$$

$$\sin 2A=2\sin A\cos A.$$

6) $\sin^2 A+\cos^2 A=E$,

$\sin(A+2\pi E)=\sin A$,

$\cos(A+2\pi E)=\cos A$,

$e^{A+i2\pi E}=e^A$.

证明

$$\sin^2 A+\cos^2 A=\left(\frac{e^{iA}-e^{-iA}}{2i}\right)^2+\left(\frac{e^{iA}+e^{-iA}}{2}\right)^2$$

$$=-\frac{e^{2iA}-2E+e^{-2iA}}{4}+\frac{e^{2iA}+2E+e^{-2iA}}{4}$$

$=E$。(注意 $e^0=E$,见 3) 的证明)

$$e^{A+i2\pi E}=e^A\cdot e^{i2\pi E}$$

$$=e^A\cdot P\begin{bmatrix}e^{2\pi i}&&&\\&e^{2\pi i}&&\\&&\ddots&\\&&&e^{2\pi i}\end{bmatrix}P^{-1}$$

$$=e^A\cdot P\cdot E\cdot P^{-1}$$

$$=e^A.$$

$$\sin(A+2\pi E)=\frac{e^{i(A+2\pi E)}-e^{-i(A+2\pi E)}}{2i}$$

$$=\frac{e^{iA}\cdot E-e^{-iA}\cdot E^{-1}}{2i}$$

$$=\frac{e^{iA}-e^{-iA}}{2i}=\sin A.$$

同理可证 $\cos(A+2\pi E)=\cos A$.

§ 8 矩阵函数在微分方程组中的应用

在线性控制系统中，常常涉及求解线性微分方程组的问题。矩阵函数在其中有着重要的应用。

我们首先讨论一阶线性常系数齐次微分方程组的定解问题：

$$
\begin{cases}
\dfrac{dX}{dt} = AX, & (1) \\
X(0) = (x_1(0), x_2(0), \cdots, x_n(0))^T. & (2)
\end{cases}
$$

这里 $A=(a_{ij})\in C^{n\times n}$，$X(t)=(x_1(t), x_2(t), \cdots, x_n(t))^T$。

设 $X(t)=(x_1(t), x_2(t), \cdots, x_n(t))^T$ 是方程组（1）的解，将 $x_i(t)$ $(i=1, 2, \cdots, n)$ 在 $t=0$ 处展开成幂级数：

$$x_i(t) = x_i(0) + x'_i(0)t + \frac{1}{2!}x''(0)t^2 + \cdots,$$

则有

$$X(t) = X(0) + X'(0)t + \frac{1}{2!}X''(0)t^2 + \cdots.$$

其中

$$X'(0) = (x'_1(0), x'_2(0), \cdots, x'_n(0))^T,$$

$$X''(0) = (x''_1(0), x''_2(0), \cdots, x''_n(0))^T,$$

$$\cdots\cdots\cdots\cdots\cdots\cdots$$

但由 $\dfrac{dX}{dt}=AX$，逐次求导可得

$$\frac{d^2X}{dt^2} = A\frac{dX}{dt} = A^2X,$$

$$\frac{d^3X}{dt^3} = \frac{d}{dt}(A^2X) = A^2\frac{dX}{dt} = A^2(AX) = A^3X,$$

$$\cdots\cdots\cdots\cdots\cdots\cdots$$

因而有

$$X'(0) = AX(0),$$

$$X''(0) = A^2X(0),$$

$X'''(0)=A^3X(0)$,

$\cdots\cdots\cdots\cdots$

所以

$$X=X(t)=X(0)+AX(0)t+\frac{1}{2!}A^2X(0)t^2+\cdots$$
$$=\left[E+(At)+\frac{1}{2!}(At)^2+\cdots\right]X(0)$$
$$=e^{At}X(0).$$

由此可见，微分方程组（1）在给定初始条件（2）下的解必定具有

$$X=e^{At}X(0)$$

的形式。下面我们来证明它确实是（1）的解。

事实上，

$$\frac{dX}{dt}=\frac{d}{dt}\left(e^{At}X(0)\right)=\left(\frac{d}{dt}e^{At}\right)X(0)+e^{At}\frac{dX(0)}{dt}$$
$$=Ae^{At}X(0)=AX.$$

又当 $t=0$ 时，$X(0)=e^0\cdot X(0)=E\cdot X(0)=X(0)$，因此这个解是满足初始条件的。

这样，我们实际上证明了下面的定理。

定理 1　一阶线性常系数微分方程组的定解问题

$$\begin{cases}\dfrac{dX}{dt}=AX,\\ X(0)=(x_1(0),x_2(0),\cdots,x_n(0))^T\end{cases}$$

有唯一解 $X=e^{At}\cdot X(0)$。

同理可证定解问题

$$\begin{cases}\dfrac{dX}{dt}=AX,\\ X|_{t=t_0}=X(t_0)\end{cases}$$

的唯一解是

$$X(t)=e^{A(t-t_0)}X(t_0).$$

最后我们考虑一阶线性常系数非齐次微分方程组的定解问题

$$\begin{cases}\dfrac{dX}{dt}=AX+F(t),\\ X|_{t=t_0}=X(t_0)。\end{cases}$$

这里 $F(t)=(F_1(t), F_2(t), \cdots, F_n(t))^T$ 是已知向量函数，A 及 X 意义同前．改写方程为

$$\frac{dX}{dt}-AX=F(t),$$

并以 e^{-At} 左乘方程两边，即得

$$e^{-At}(\frac{dX}{dt}-AX)=e^{-At}F(t),$$

即

$$\frac{d}{dt}(e^{-At}X)=e^{-At}F(t)。$$

在 $[t_0, t]$ 上进行积分，可得

$$e^{-At}X-e^{-At_0}X(t_0)=\int_{t_0}^{t}e^{-A\tau}F(\tau)d\tau$$

即

$$X=e^{A(t-t_0)}X(t_0)+\int_{t_0}^{t}e^{A(t-\tau)}F(\tau)d\tau.$$

它就是我们考虑的定解问题的解．

例 1 求定解问题

$$\begin{cases}\dfrac{dX}{dt}=AX\\ X(0)=(1,1,1)^T\end{cases}\qquad A=\begin{bmatrix}3&-1&1\\2&0&-1\\1&-1&2\end{bmatrix}$$

的解

$$|\lambda E-A|=\begin{vmatrix}\lambda-3&1&-1\\-2&\lambda&1\\-1&1&\lambda-2\end{vmatrix}=\lambda(\lambda-2)(\lambda-3),$$

故 A 有三个不同的特征值．从而 A 可与对角形矩阵相似．与特征

值 $\lambda_1=0$，$\lambda_2=2$，$\lambda_3=3$ 相应的三个线性无关的特征向量为：

$$X_1=(1,5,2)^T, X_2=(1,1,0)^T, X_3=(2,1,1)^T.$$

故得

$$P=\begin{bmatrix}1 & 1 & 2\\5 & 1 & 1\\2 & 0 & 1\end{bmatrix} \text{及 } P^{-1}=-\frac{1}{6}\begin{bmatrix}1 & -1 & -1\\-3 & -3 & 9\\-2 & 2 & -4\end{bmatrix}$$

所以，由定理 1 可得所求的解为

$$X=e^{At}\cdot X(0)=P\cdot\begin{bmatrix}1 & & \\ & e^{2t} & \\ & & e^{3t}\end{bmatrix}\cdot P^{-1}\cdot X(0)$$

$$=\begin{bmatrix}1 & 1 & 2\\5 & 1 & 1\\2 & 0 & 1\end{bmatrix}\cdot\begin{bmatrix}1 & & \\ & e^{2t} & \\ & & e^{3t}\end{bmatrix}\cdot\left(-\frac{1}{6}\right)\begin{bmatrix}1 & -1 & -1\\-3 & -3 & 9\\-2 & 2 & -4\end{bmatrix}\cdot\begin{bmatrix}1\\1\\1\end{bmatrix}$$

$$=-\frac{1}{6}\begin{bmatrix}-1+3e^{2t}-8e^{3t}\\-5+3e^{2t}-4e^{3t}\\-2-4e^{3t}\end{bmatrix}$$

例 2 求定解问题

$$\begin{cases}\dfrac{dX}{dt}=AX+F(t)\\X(0)=(1,1,1)^T\end{cases}$$

的解，其中矩阵 A 与例 1 的 A 相同，又 $F(t)=(0, 0, e^{2t})^T$。

解 由前面的讨论，这问题的解为

$$X=e^{At}X(0)+\int_0^t e^{A(t-\tau)}F(\tau)d\tau$$

这里 $e^{At}X(0)$ 在例 1 已经求出，故只须计算积分

$$I=\int_0^t e^{A(t-\tau)}F(\tau)d\tau$$

由于（P 与 P^{-1} 见例 1）：

$$e^{A(t-\tau)}F(\tau)=P\cdot\begin{bmatrix}1 & & \\ & e^{2(t-\tau)} & \\ & & e^{3(t-\tau)}\end{bmatrix}P^{-1}\begin{bmatrix}0\\0\\e^{2\tau}\end{bmatrix}$$

$$=P\cdot\begin{bmatrix}1 & & \\ & e^{2(t-\tau)} & \\ & & e^{3(t-\tau)}\end{bmatrix}\left(-\frac{1}{6}\right)\begin{bmatrix}-e^{2\tau}\\9e^{2\tau}\\-4e^{2\tau}\end{bmatrix}$$

$$=-\frac{1}{6}P\cdot\begin{bmatrix}-e^{2\tau}\\9e^{2t}\\-4e^{3t-\tau}\end{bmatrix}$$

$$=-\frac{1}{6}\begin{bmatrix}-e^{2\tau}+9e^{2t}-8e^{3t-\tau}\\-5e^{2\tau}+9e^{2t}-4e^{3t-\tau}\\-e^{2\tau}-4e^{3t-\tau}\end{bmatrix}$$

将这一结果对变量 τ 从 0 到 t 进行积分，即得

$$I=-\frac{1}{6}\begin{bmatrix}\frac{1}{2}+(9t+\frac{15}{2})e^{2t}-8e^{3t}\\\frac{5}{2}+(9t+\frac{3}{2})e^{2t}-4e^{3t}\\1+3e^{2t}-4e^{3t}\end{bmatrix}$$

因此 $X=e^{At}X(0)+I$，

即

$$X=-\frac{1}{6}\begin{bmatrix}-\frac{1}{2}+(9t+\frac{21}{2})e^{2t}-16e^{3t}\\-\frac{5}{2}+(9t+\frac{9}{2})e^{2t}-8e^{3t}\\-1+3e^{2t}-8e^{3t}\end{bmatrix}$$

以上我们仅仅介绍了矩阵函数在一阶线性常系数齐次及非齐次微分方程组中的应用，这是较为简单的情形。同样地，在一阶线性变系数齐次及非齐次微分方程组中，矩阵函数都有应用，但都已经远比上面所讲的困难多了，在此不拟作更深入的讨论。

§9 线性系统的能控性与能观测性

作为矩阵函数的一个应用，这里简单讨论一下现代控制论中两个基本概念：系统的能控性与能观测性。我们只就连续型的线性定常系统进行讨论。这系统为

$$\begin{cases}\dfrac{\mathrm{d}X(t)}{\mathrm{d}t}=AX(t)+Bu(t), & (1)\\ Y(t)=CX(t)+Du(t) & (2)\end{cases}$$

其中 A，B，C，D 均为常数矩阵。系统矩阵 A 是 $n\times n$ 矩阵，输入矩阵 B 是 $n\times m$ 的、输出矩阵 C 是 $p\times n$ 的，又矩阵 D 是 $p\times m$ 的。状态向量 $X(t)$ 是 n 维列向量，输入向量 $u(t)$ 与输出向量 $Y(t)$ 分别是 m 维、p 维列向量。这个系统简称为系统 (A, B, C)。

定义 1 对于一个线性定常系统，若在某个有限时间区 $[0, t_1]$ 内存在着输入 $u(t)$ $(0\leqslant t\leqslant t_1)$，能使系统从任意初始状态 $X(0)=X_0$ 转移到 $X(t_1)=\theta$，则称此状态 X_0 是能控的；若系统的所有状态都是能控的，则称此系统是完全能控的。

定理 1 系统 (A, B, C) 完全能控的充要条件，是 n 阶对称矩阵

$$W_c(0,t_1)=\int_0^{t_1}\mathrm{e}^{-A\tau}BB^T\mathrm{e}^{-A^T\tau}\mathrm{d}\tau \tag{3}$$

为非奇异矩阵。

证明 充分性。设 $W_c(0, t_1)$ 非奇异。对方程 (1) 从 0 到 t_1 积分得

$$X(t_1)=\mathrm{e}^{At_1}X(0)+\int_0^{t_1}\mathrm{e}^{A(t_1-\tau)}Bu(\tau)\mathrm{d}\tau, \tag{4}$$

令

$$u(t)=-B^T\mathrm{e}^{-A^Tt}W_c^{-1}(0,t_1)X(0), \tag{5}$$

把 $u(t)$ 代入 (4) 式得

$$X(t_1)=e^{At_1}X(0)-e^{At_1}\left(\int_0^{t_1}e^{-A\tau}BB^Te^{-A^T\tau}d\tau\right)W_c^{-1}(0,t_1)X(0)$$
$$=e^{At_1}X(0)-e^{At_1}W_c(0,t_1)W_c^{-1}(0,t_1)X(0)$$
$$=e^{At_1}X(0)-e^{At_1}X(0)$$
$$=\theta.$$

这说明在（5）式所示的控制输入 $u(t)$ 作用下，能使系统从 $X(0)$ 转移到 $X(t_1)=\theta$。

必要性。用反证法。若系统是完全能控的，但 $W_c(0, t_1)$ 是奇异的，则将引出矛盾。

因 $W_c(0, t_1)$ 是奇异的，则必有非零向量 $\alpha=(a_1, a_2, \cdots, a_n)^T$，使对任意时刻 $t=t_1\geqslant 0$，下式能够成立：

$$\alpha^TW_c(0,t_1)\alpha=0,$$

即
$$\int_0^{t_1}\alpha^T(e^{-A\tau}BB^Te^{-A^T\tau})\alpha d\tau=0.$$

故对任意时刻 t，有

$$\alpha^Te^{-At}B=0^T\quad(t\geqslant 0).\tag{6}$$

现系统是完全能控的，故由前面已证明的充分条件，必存在某个 $u(t)$，使其作用于系统上，使得 $X(t_1)=\theta$，故由（4）式得

$$-X(0)=\int_0^{t_1}e^{-A\tau}Bu(\tau)d\tau.$$

上式两边左乘以 α^T，并考虑到（6）式，便有

$$-\alpha^TX(0)=\int_0^{t_1}\alpha^Te^{-A\tau}Bu(\tau)d\tau=0.$$

由于 $X(0)$ 为任意的，现选 $X(0)=\alpha$，则由上式得

$$\alpha^T\alpha=0.$$

这与 α 是非零向量相矛盾，故 $W_c(0, t_1)$ 是非奇异的。证毕。

用定理 1 来判别系统的能控制性是不方便的，但在理论上有其重要性。下面的定理提供了一个较简便的判别准则。

定理 2．系统 (A, B, C) 完全能控的充要条件，是 $n\times mn$ 矩

阵

$$W_c = [B, AB, A^2B, \cdots, A^{n-1}B]$$

的秩为 n。W_c 称为能控性矩阵。

这定理可用反证法来证明。这里只给出充分性的证明。如果系统不是完全能控的，则由定理 1 知矩阵 $W_c(0, t_1)$ 是奇异的，故由上面的（6）式两边求 k 次微商可得：

$$\alpha^T(-A)^k e^{-At}B = \theta \quad (k = 1, 2, \cdots, n-1)。$$

令 $t=0$ 得

$$\alpha^T A^k B = \theta \quad (k = 1, 2, \cdots, n-1),$$

而在（6）式中令 $t=0$，得 $\alpha^T B=\theta$，故对于 $k=0$，1，2，…，$n-1$ 均有

$$\alpha^T A^k B = \theta。$$

即 α^T $[B, AB, A^2B, \cdots, A^{n-1}B]=\theta$（$\alpha$ 是非零向量）。这与矩阵 W_c 的秩为 n 相矛盾，就证明了系统是完全能控的。

例 1 已知

$$A = \begin{bmatrix} 1 & 3 & 2 \\ 0 & 2 & 0 \\ 0 & 1 & 3 \end{bmatrix} \qquad B = \begin{bmatrix} 2 & 1 \\ 1 & 1 \\ -1 & -1 \end{bmatrix}$$

试判别系统（A，B，C）是否完全能控。

解 这是一个具有两个输入，三个状态变量的系统，现用定理 2 来判定其能控性。因为

$$W_c = [B, AB, A^2B] = \begin{bmatrix} 2 & 1 & 3 & 2 & 5 & 4 \\ 1 & 1 & 2 & 2 & 4 & 4 \\ -1 & -1 & -2 & -2 & -4 & -4 \end{bmatrix}$$

$$\xrightarrow{\text{初等变换}} \begin{bmatrix} 2 & 1 & 3 & 2 & 5 & 4 \\ 1 & 1 & 2 & 2 & 4 & 4 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

故 W_c 的秩 $\neq 3$，从而系统不是完全能控的。

下面我们来考虑能观测性问题。

定义 2 对于一个线性定常系统，若在有限时间区间 $[0, t_1]$ 内，能通过观测系统的输出 $Y(t)$ 而唯一地确定任意初始状态 $X(0)$，则称此系统是完全能观测的，或者说对每一状态 $X(0)$ 是能观测的。

定理 3 系统 (A, B, C) 完全能观测的充要条件，是 n 阶对称矩阵

$$M(0, t_1) = \int_0^{t_1} e^{A^T \tau} C^T C e^{A\tau} d\tau$$

为非奇异矩阵。

证明 充分性：

$$\begin{aligned} Y(t) &= CX(t) + Du(t) \\ &= Ce^{At}X(0) + C\int_0^t e^{A(t-\tau)}Bu(\tau)d\tau + Du(t). \end{aligned}$$

令 $$\eta(t) = Y(t) - C\int_0^t e^{A(t-\tau)}Bu(\tau)d\tau - Du(t),$$

于是得

$$Ce^{At}X(0) = \eta(t). \qquad (*)$$

以 $e^{A^T t} \cdot C^T$ 左乘上式两边，并从 0 到 t_1 进行积分，可得

$$\left(\int_0^{t_1} e^{A^T t} C^T C e^{At} dt\right) X(0) = \int_0^{t_1} e^{A^T t} C^T \eta(t) dt$$

即 $$M(0, t_1) X(0) = \int_0^{t_1} e^{A^T t} C^T \eta(t) dt.$$

若 $M(0, t_1)$ 为非奇异矩阵，则

$$X(0) = M^{-1}(0, t_1) \int_0^{t_1} e^{A^T t} C^T \eta(t) dt,$$

即 $X(0)$ 能唯一确定。由此可知，若 $M(0, t_1)$ $(t_1 > 0)$ 为非奇异，则系统是完全能观测的。

必要性：

若系统是完全能观测的．现用反证法证明 $M(0, t_1)$ 是非奇

异的。事实上，如果 $M(0,t_1)$ 是奇异的，则存在非零 n 维向量 α，使得对任意 $t\geqslant 0$，有

$$\alpha^T M(0,t_1)\alpha = 0 .$$

将 $M(0,t_1)$ 代入此式便得

$$\alpha^T \left(\int_0^{t_1} e^{A^T\tau} C^T C e^{A\tau} d\tau \right)\alpha = 0 .$$

这表明对任意 $t\geqslant 0$ 时，有

$$Ce^{At}\alpha = \theta。$$

若取 $\alpha = X(0) \neq \theta$，则对任意 $t\geqslant 0$，便有

$$Ce^{At}X(0) = \theta \quad (X(0) \neq \theta) .$$

这与当 $X(0)=\theta$ 时由（*）式

$$\eta(t) = Ce^{At}X(0) = \theta \quad (X(0) = \theta)$$

的结果作比较，说明 $X(0)$ 不能唯一确定。而这与系统是完全能观测的假设相矛盾。定理证毕。

定理 4 系统 (A, B, C) 完全能观测的充要条件，是 $P^{\times n}$ 矩阵

$$W_0 = \begin{pmatrix} C \\ CA \\ \vdots \\ CA^{n-1} \end{pmatrix}$$

的秩等于 n，亦即 $n\times pn$ 矩阵

$$W_0^T = [C^T, A^T C^T, \cdots, (A^T)^{n-1} C^T]$$

的秩等于 n。矩阵 W_0 叫做能观测性矩阵。

例 2 若某系统的状态方程与输出方程为

$$\frac{dX(t)}{dt} = \begin{pmatrix} -4 & 1 \\ -6 & 1 \end{pmatrix} X(t) + \begin{pmatrix} 3 \\ 7 \end{pmatrix} u(t)$$

$$Y(t) = \begin{pmatrix} 1 & 1 \\ 2 & 3 \end{pmatrix} X(t)$$

试判定这系统的能观测性。

解 它的能观测性矩阵为

$$W_0=\begin{bmatrix}C\\CA\end{bmatrix}=\begin{bmatrix}1&1\\2&3\\-10&2\\-26&5\end{bmatrix}$$

显然它的秩等于 2，故系统是能观测的。

习 题 四

1. 若 $\|\cdot\|$ 是酉空间 C^n 的向量范数，证明向量范数的下列基本性质：

(1) 零向量的范数是零；

(2) 当 x 是非零向量，则

$$\left\|\frac{x}{\|x\|}\right\|=1;$$

(3) $\|-x\|=\|x\|$；

(4) $|\|x\|-\|y\||\leqslant\|x-y\|$.

2. 证明：若 $x\in C^n$，则

(1) $\|x\|_2\leqslant\|x\|_1\leqslant\sqrt{n}\|x\|_2$；

(2) $\|x\|_\infty\leqslant\|x\|_1\leqslant n\|x\|_\infty$；

(3) $\|x\|_\infty\leqslant\|x\|_2\leqslant\sqrt{n}\|x\|_\infty$.

3. 证明：在 R^n 中当且只当 x，y 线性相关而且 $x^Ty\geqslant 0$ 时，才有 $(x, y)=\|x\|_2\cdot\|y\|_2$.

4. 若 T 为正交矩阵，又 $A\in R^{n\times n}$，证明：

(1) $\|T\|_2=1$；

(2) $\|A\|_2=\|TA\|_2$.

5. 证明：

$$\frac{1}{\sqrt{n}}\|A\|_F\leqslant\|A\|_2\leqslant\|A\|_F.$$

6. 设

$$A=\begin{pmatrix}1 & 2\\ 2 & 1\end{pmatrix}$$

试求 e^A。

7. 对下列方阵 A，求矩阵函数 e^{At}：

(1) $A=\begin{pmatrix}0 & 1\\ -2 & -3\end{pmatrix}$

(2) $A=\begin{bmatrix}2 & -2 & 3\\ 1 & 1 & 1\\ 1 & 3 & -1\end{bmatrix}$

(3) $A=\begin{bmatrix}0 & 1 & 0\\ 0 & 0 & 1\\ -8 & -12 & -6\end{bmatrix}$

(4) $A=\begin{bmatrix}-2 & 1 & 3\\ 0 & -3 & 0\\ 0 & 2 & -2\end{bmatrix}$

8. 设

$$A=\begin{pmatrix}0 & 1\\ 0 & -2\end{pmatrix}$$

求 e^A， $\sin A$， $\cos At$。

9. 设

$$A(t)=\begin{pmatrix}\cos t & \sin t\\ -\sin t & \cos t\end{pmatrix}$$

求 $\frac{d}{dt}A(t)$， $\frac{d}{dt}A^{-1}(t)$， $\frac{d}{dt}|A(t)|$， $\left|\frac{d}{dt}A(t)\right|$。

10. 设

$$A(t)=\begin{bmatrix}e^{2t} & te^{t} & 1\\ e^{-t} & 2e^{2t} & 0\\ 3t & 0 & 0\end{bmatrix}$$

求 $\int A(t)dt$， $\int_0^1 A(t)dt$。

11. 证明：

(1) 若 A 为实反对称阵，则 e^A 为正交阵；

(2) 若 A 为厄米特矩阵，则 e^{iA} 为酉矩阵。

12. 求微分方程的解：

$$\begin{cases} \dfrac{dX}{dt} = \begin{pmatrix} -1 & 2 \\ -2 & 1 \end{pmatrix} X(t) \\ X(0) = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \end{cases}$$

13. **求非齐次微分方程的解：**

$$\begin{cases} \dfrac{dX}{dt} = \begin{pmatrix} 3 & 5 \\ -5 & 3 \end{pmatrix} X(t) + \begin{pmatrix} e^{-t} \\ 0 \end{pmatrix} \\ X(0) = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \end{cases}$$

第五章　特征值的估计与广义逆矩阵

特征值的估计与广义逆矩阵是矩阵理论中两个不同的专门课题，两者都有丰富的内容和许多重要的应用。在本章里，我们仅就这两个方面的内容作一个基本的概述。

矩阵特征值的计算与估计在理论上和实际应用中都是重要的，但要精确计算特征值并非总是可能的，即使在某些特殊情况下有可能，可是付出的代价也是很大的。幸好在许多应用中并不需要精确计算矩阵的特征值，而只需有一个粗略的估计就够了，例如：在线性系统理论中，通过估计系统矩阵 A 的特征值是否有负实部，便可判定系统的稳定性；当研究一个迭代法的收敛性时便要判断迭代矩阵的特征值是否都落在单位圆内；在差分方法的稳定性理论以及自动控制理论中都需要估计矩阵的特征值是否在复数平面上的某一确定的区域中。

本章要讨论的另一类问题是广义逆矩阵方面的问题。我们知道，若方阵 A 的行列式不等于零，则存在唯一的方阵 B，满足 $AB=BA=E$，并称 B 为 A 的逆矩阵，记为 A^{-1}。当 A 不是方阵，或方阵 A 的行列式等于零时，则上述的逆矩阵就不存在。然而，Moore 在 1920 年便将逆矩阵的概念推广到任意矩阵上，他是用正交投影算子来定义广义逆矩阵的，人们把他定义的广义逆矩阵称为 Moore 广义逆。1955 年，Penrose 用方程组

$$AGA = A, GAG = G, (AG)^H = AG, (GA)^H = GA$$

来定义 A 的广义逆。不久以后，Bjerhammer 证明了 Moore 逆与 Penrose 逆的等价性，所以后来把它叫做 Moore－Penrose 逆，并记为 A^+。此后，对广义逆矩阵的研究又有很大的发展，现已形成一套系

统的理论，且在系统分析、优化计算及统计学等领域中有许多的应用。这里主要介绍15种广义逆矩阵中较常用的 A^- 及 A^+ 两种，其他就不一一介绍了。

§1 特征值的界的估计

设 $A=(a_{ij})_{n\times n}$ 为一给定的复数矩阵，则 A 可以表示成一个厄米特矩阵 B 与一个反厄米特矩阵 C 之和，事实上，B,C 便是如下的矩阵：

$$B=(b_{ij})_{n\times n}=\frac{A+A^H}{2}\quad (b_{ij}=\frac{a_{ij}+\bar{a}_{ji}}{2}),$$

$$C=(c_{ij})_{n\times n}=\frac{A-A^H}{2}\quad (c_{ij}=\frac{a_{ij}-\bar{a}_{ji}}{2})。$$

设 A,B,C 的特征值的集合分别为：

$$\{\lambda_1,\lambda_2,\cdots,\lambda_n\},\{\mu_1,\mu_2,\cdots,\mu_n\},\{iv_1,iv_2,\cdots,iv_n\}.$$

这里的每个 μ_j 及 v_j 都是实数。并假设

$$|\lambda_1|\geqslant|\lambda_2|\geqslant\cdots\geqslant|\lambda_n|,\mu_1\geqslant\mu_2\geqslant\cdots\geqslant\mu_n,\quad v_1\geqslant v_2\geqslant\cdots\geqslant v_n.$$

定理1 若 n 阶复数矩阵 $A=(a_{ij})$ 的特征值的集合（A 的谱）为 $\{\lambda_1,\lambda_2,\cdots,\lambda_n\}$，则有不等式

$$\sum_{i=1}^{n}|\lambda_i|^2\leqslant\sum_{i=1}^{n}\sum_{j=1}^{n}|a_{ij}|^2,$$

又等号当且只当 A 为正规矩阵时成立。

证明 由第三章的舒尔定理，存在酉矩阵 U 及上三角矩阵 T，使得

$$U^HAU=T,$$

因此有 $U^HA^HU=T^H$，从而得

$$U^HAA^HU=TT^H.$$

$$\mathrm{tr}(AA^H)=\mathrm{tr}(U^HAA^HU)=\mathrm{tr}(TT^H).\qquad(1)$$

由于矩阵 T 的主对角线上的元素都是 A 的特征值，故由(1)式得

$$\sum_{i=1}^{n}|\lambda_i|^2=\sum_{i=1}^{n}|t_{ii}|^2\leqslant\sum_{i=1}^{n}\sum_{j=1}^{n}|t_{ij}|^2。\qquad(2)$$

而(2)式右端是矩阵 T 的 Frobenius 范数的平方。由于 A 与 T 是酉相似的($U^HAU=T$)，而在酉相似下，矩阵的 Frobenius 范数不变。所以

$$\sum_{i=1}^{n}\sum_{j=1}^{n}|t_{ij}|^2=\sum_{i=1}^{n}\sum_{j=1}^{n}|a_{ij}|^2。\qquad(3)$$

综合(2)、(3)两式便得到所需证的不等式。又不等式(2)取等号当且只当 $i\neq j$ 时都有 $t_{ij}=0$，即 A 酉相似于对角形矩阵，也就是 A 为正规矩阵(见第二章§7)。证毕。

推论 1 若 A,B,C 如前所设，则有

(1) $|\lambda_i|\leqslant n\cdot\max\limits_{1\leqslant i,j\leqslant n}|a_{ij}|$，

(2) $|\mathrm{Re}(\lambda_i)|\leqslant n\cdot\max\limits_{1\leqslant i,j\leqslant n}|b_{ij}|$，

(3) $|\mathrm{Im}(\lambda_i)|\leqslant n\cdot\max\limits_{1\leqslant i,j\leqslant n}|c_{ij}|$。

证明 由定理 1 的证明中所得到的等式

$$U^HAU=T,\ U^HA^HU=T^H,$$

以及 B,C 的定义，可得

$$U^HBU=U^H(\frac{A+A^H}{2})U=\frac{1}{2}(T+T^H),$$

$$U^HCU=U^H(\frac{A-A^H}{2})U=\frac{1}{2}(T-T^H).$$

注意到 T 是上三角矩阵，且 T 的主对角线上的元素都是 A 的特征值，又在酉相似下矩阵的 Frobenius 范数不变，所以有

$$\sum_{i=1}^{n}\left|\frac{\lambda_i+\bar{\lambda}_i}{2}\right|^2+\sum_{i\neq j}\frac{1}{2}|t_{ij}|^2=\sum_{1\leqslant i,j\leqslant n}|b_{ij}|^2\leqslant n^2\cdot\max_{1\leqslant i,j\leqslant n}|b_{ij}|^2,$$

$$\sum_{i=1}^{n}\left|\frac{\lambda_i-\bar{\lambda}_i}{2}\right|^2+\sum_{i\neq j}\frac{1}{2}|t_{ij}|^2=\sum_{1\leqslant i,j\leqslant n}|c_{ij}|^2\leqslant n^2\cdot\max_{1\leqslant i,j\leqslant n}|c_{ij}|^2.$$

于是有

$$\sum_{i=1}^{n}|\mathrm{Re}(\lambda_i)|^2=\sum_{i=1}^{n}\left|\frac{\lambda_i+\bar{\lambda}_i}{2}\right|^2\leqslant n^2\cdot\max_{1\leqslant i,j\leqslant n}|b_{ij}|^2,$$

$$\sum_{i=1}^{n}|\mathrm{Im}(\lambda_i)|^2=\sum_{i=1}^{n}\left|\frac{\lambda_i-\bar{\lambda}_i}{2}\right|^2\leqslant n^2\cdot\max_{1\leqslant i,j\leqslant n}|c_{ij}|^2.$$

当然对任一 $i\in\{1,2,\cdots,n\}$ 都有

$$|\mathrm{Re}(\lambda_i)|^2\leqslant n^2\cdot\max_{1\leqslant i,j\leqslant n}|b_{ij}|^2,$$

$$|\mathrm{Im}(\lambda_i)|^2\leqslant n^2\cdot\max_{1\leqslant i,j\leqslant n}|c_{ij}|^2.$$

由此即得(2)式及(3)式。再由定理 1 得

$$\sum_{i=1}^{n}|\lambda_i|^2\leqslant\sum_{1\leqslant i,j\leqslant n}|a_{ij}|^2\leqslant n^2\cdot\max_{1\leqslant i,j\leqslant n}|a_{ij}|^2,$$

因而有

$$|\lambda_i|\leqslant n\cdot\max_{1\leqslant i,j\leqslant n}|a_{ij}|,$$

即(1)式成立。证毕。

由推论 1 可以证明(从略)下面的推论：

推论 2 设 $A=(a_{ij})$ 是 n 阶实矩阵，则

$$|\mathrm{Im}(\lambda_i)|\leqslant\sqrt{\frac{n(n-1)}{2}}\max_{1\leqslant i,j\leqslant n}|c_{ij}|,$$

这里

$$c_{ij}=\frac{1}{2}(a_{ij}-\bar{a}_{ji})\quad i,j=1,2,\cdots,n.$$

例 1 估计下面矩阵的特征值的界限：

$$A=\begin{bmatrix}0 & 0.2 & 0.1\\ -0.2 & 0 & 0.2\\ -0.1 & -0.2 & 0\end{bmatrix}$$

解 因为

$$B=\frac{1}{2}(A+A^H)=0(\text{矩阵}),$$

$$C=\frac{1}{2}(A-A^H)=A。$$

所以由推论 1 即有

$$|\lambda_i|\leqslant 3\times 0.2=0.6;$$

$$|\mathrm{Re}(\lambda_i)|\leqslant 3\times 0=0,\quad \text{即}\ \mathrm{Re}(\lambda_i)=0;$$

$$|\mathrm{Im}(\lambda_i)|\leqslant 3\times 0.2=0.6。$$

若应用推论 2 即有

$$|\mathrm{Im}(\lambda_i)|\leqslant\sqrt{\frac{3\times(3-1)}{2}}\times 0.2=0.3464。$$

故所给矩阵 A(实反对称矩阵)的特征值的模不超过 0.3464,且由推论 2 得到的估计值要比由推论 1 得到的更好些(由直接计算可求出 A 的特征值为 $\lambda_1=0,\lambda_2=-0.3i,\lambda_3=0.3i$)。

§2 圆盘定理

上节我们对矩阵特征值作了大致的估计。本节将对矩阵的特征值在复数平面上的位置作更准确的估计.这就是圆盘定理(又称 Gerschgorin 定理)所表述的。圆盘定理有很多应用和推广,但在这里不打算作详细讨论。

下面的两个定理都称为圆盘定理。

定理 1 设 $A=(a_{ij})$为任一 n 阶复数矩阵,则 A 的特征值都在复数平面上的 n 个圆盘

$$|z-a_{ii}|\leqslant R_i\quad(i=1,2,\cdots,n)$$

的并集内;又这里的

$$R_i=|a_{i1}|+|a_{i2}|+\cdots+|a_{i,i-1}|+|a_{i,i+1}|+\cdots+|a_{in}|。$$

证明 设 λ 为 A 的特征值,其对应的一个特征向量为 $x(\neq$

θ),即 $Ax=\lambda x$,写成分量形式就是:

$$\sum_{j=1}^{n} a_{ij}x_j = \lambda x_i \quad (i=1,2,\cdots,n),$$

或

$$\sum_{j\neq i} a_{ij}x_j = (\lambda - a_{ii})x_i \quad (i=1,2,\cdots,n)。$$

设 x_t 为 x 的各分量中绝对值最大的一个,则 $x_t\neq 0$。对等式

$$(\lambda - a_{tt})x_t = \sum_{j\neq t} a_{tj}x_j$$

两边除以 x_t 并取绝对值便得

$$|\lambda - a_{tt}| \leqslant \sum_{j\neq t} |a_{tj}| \cdot \left|\frac{x_j}{x_t}\right| \leqslant R_t,$$

即 λ 在圆盘

$$|z - a_{tt}| \leqslant R_t$$

中,当然也在 n 个圆盘 $|z-a_{ii}|\leqslant R_i \quad (i=1,2,\cdots,n)$ 的并集之中。证毕。

上述的圆盘 $|z-a_{ii}|\leqslant R_i$ 称为 Gerschgorin 圆盘,简称盖尔圆。定理 1 表明对于 A 的任一特征值 λ,总存在盖尔圆 S_i,使得 $\lambda\in S_i$。

例 1 估计下面矩阵的特征值范围:

$$A = \begin{bmatrix} 1 & 0.1 & 0.2 & 0.3 \\ 0.5 & 3 & 0.1 & 0.2 \\ 1 & 0.3 & -1 & 0.5 \\ 0.2 & -0.3 & -0.1 & -4 \end{bmatrix}$$

解 圆盘定理 1 所指的四个圆盘为:

$$|z-1| \leqslant 0.1+0.2+0.3=0.6,$$
$$|z-3| \leqslant 0.5+0.1+0.2=0.8,$$
$$|z+1| \leqslant 1+0.3+0.5=1.8,$$
$$|z+4| \leqslant 0.2+0.3+0.1=0.6。$$

而 A 的特征值都落在这四个盖尔圆的并集内,读者可在复数平面上画出相应的图形,从图中可以看到,第一、第三这两个圆盘是相

交的，而另外的两个都是孤立的(即除自身外不与其他三个圆盘的任何一个相交)。两个相交的圆盘的并集构成一个连通区域。一般地说，由矩阵 A 的 k 个相交的盖尔圆的并集构成的连通区域称为一个连通部分，并说它是由 k 个盖尔圆组成的。

定理 1 只是说明矩阵 A 的特征值落在 A 的 n 个盖尔圆的并集中，并无指出在那个圆盘中有多少个特征值。下述定理更深刻地表述了 A 的特征值的分布状况。

定理 2 矩阵 A 的任一由 k 个盖尔圆组成的连通部分里，有且只有 A 的 k 个特征值(当 A 的主对角线上有相同元素时，则按重复次数计算，有相同特征值时亦需按重复次数计算)。

这定理的证明颇为复杂，这里就不作介绍了。从这个定理可以得知，由一个盖尔圆组成的连通部分，有且只有一个特征值；由两个盖尔圆组成的连通部分有且只有两个特征值，但有可能这两个特征值都落在两个圆盘的一个之中，而不在另一个之中(见下例)。

例 2 矩阵

$$A = \begin{bmatrix} 1 & -0.8 \\ 0.5 & 0 \end{bmatrix}$$

的特征方程为

$$(\lambda - 1)\lambda + 0.4 = 0.$$

故 A 的特征值为

$$\lambda_1 = \frac{1 + \sqrt{0.6}i}{2}, \quad \lambda_2 = \frac{1 - \sqrt{0.6}i}{2}.$$

又 A 的两个盖尔圆盘为

$$|z - 1| \leqslant 0.8, \quad |z| \leqslant 0.5,$$

但因 λ_i 的模

$$|\lambda_1| = |\lambda_2| = \sqrt{0.4} = 0.632456 > 0.5.$$

因此，这两个特征值都不落在圆盘 $|z| \leqslant 0.5$ 之中。

§3 谱半径的估计

我们在第四章研究矩阵幂级数收敛性时，已定义过矩阵 A 的谱半径 $\rho(A)$，并通过不等式 $\rho(A)<R$ 刻画了矩阵 A 的幂级数的收敛性。矩阵的谱半径及其估计在线性方程组的迭代法收敛性问题上，以及在差分方程组的稳定性问题上，都有重要的应用。

我们已经知道，若 $A=(a_{ij})$ 是复数域上的 n 阶方阵，又 $\lambda_1,\lambda_2,\cdots,\lambda_n$ 是 A 的全部特征值，则

$$\rho(A)=\max_{1\leqslant i\leqslant n}|\lambda_i|$$

称为 A 的谱半径。

下面我们对 $\rho(A)$ 作一些估计。

在讨论矩阵的范数时，我们曾经证明过下面的结果：

矩阵 A 的每一个特征值的模（绝对值），都不超过矩阵 A（在任意一种矩阵范数 $\|\cdot\|$ 定义下）的范数 $\|A\|$，即 $|\lambda_i|\leqslant\|A\|$。

由此即得：

定理1 复数域上的任一 n 阶方阵 $A=(a_{ij})$ 的谱半径 $\rho(A)$ 都不超过 A 的范数 $\|A\|$，即

$$\rho(A)\leqslant\|A\|,$$

这里 $\|A\|$ 是在任意一种方阵范数 $\|\cdot\|$ 定义下方阵 A 的范数。

若取方阵范数 $\|\cdot\|$ 为 $\|\cdot\|_1$，$\|\cdot\|_\infty$ 或 $\|\cdot\|_2$，则有下面的推论：

推论 1) $\rho(A)\leqslant\|A\|_1=\max\limits_{1\leqslant j\leqslant n}\sum\limits_{i=1}^{n}|a_{ij}|$；

2) $\rho(A)\leqslant\|A\|_\infty=\max\limits_{1\leqslant i\leqslant n}\sum\limits_{j=1}^{n}|a_{ij}|$；

3) $\rho(A)\leqslant\|A\|_2=\sqrt{\lambda_{A^HA}}$，

这里 λ_{A^HA} 为矩阵 A^HA 的最大特征值。

当 A 是正规矩阵时，则有下述定理。

定理 2 若 A 为 n 阶正规矩阵，则

$$\rho(A) = \| A \|_2 .$$

证明 因 A 是正规矩阵，故存在酉矩阵 P，使得

$$P^HAP = \begin{pmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{pmatrix}$$

由此易得

$$P^HA^HP = \begin{pmatrix} \bar{\lambda}_1 & & & \\ & \bar{\lambda}_2 & & \\ & & \ddots & \\ & & & \bar{\lambda}_n \end{pmatrix}.$$

从而

$$P^HA^HAP = \begin{pmatrix} |\lambda_1|^2 & & & \\ & |\lambda_2|^2 & & \\ & & \ddots & \\ & & & |\lambda_n|^2 \end{pmatrix}$$

又显然有

$$\lambda_{A^HA} = \max\{|\lambda_1|^2, |\lambda_2|^2, \cdots, |\lambda_n|^2\} = |\lambda_t|^2,$$

这里 t 是 $\{1,2,\cdots,n\}$ 中的某个值。因此有

$$\| A \|_2 = \sqrt{\lambda_{A^HA}} = |\lambda_t|.$$

而

$$\rho(A) = \max\{|\lambda_1|, |\lambda_2|, \cdots, |\lambda_n|\} = |\lambda_t|,$$

所以 $\rho(A) = \| A \|_2$。证毕。

由于对角形矩阵、实对称矩阵、实反对称矩阵、正交矩阵、酉矩阵、厄米特矩阵、反厄米特矩阵都是正规矩阵，所以对于它们都具有性质 $\rho(A) = \| A \|_2$。

§4 广义逆矩阵与线性方程组的解

设有线性方程组

$$Ax = b, \tag{1}$$

这里 A 和 b 分别是 $m\times n$ 和 $m\times 1$ 矩阵*。若 $m=n$ 且 A 可逆,则这方程组对任何 b 都有唯一的解 $x=A^{-1}b$。当 $m\neq n$,或 $m=n$ 但 A 不可逆,方程组都不一定有解;即使有解,解也不一定唯一。如果知道方程组(1)有解,我们问:可否用类似于 A 可逆时解的表示式 $x=A^{-1}b$ 那样来表示方程组的解?

定义 1 设 A 是 $m\times n$ 矩阵。一个 $n\times m$ 矩阵 G 称为 A 的一个 $\{1\}$-广义逆,如果对任意给出的 $m\times 1$ 矩阵 b,只要方程组 $Ax=b$ 有解,则 $x=Gb$ 也一定是解。

定理 1 $n\times m$ 矩阵 G 是 $m\times n$ 矩阵 A 的一个$\{1\}$-广义逆,当且只当 $AGA=A$。

证明 充分性。设 x 是方程组 $Ax=b$ 的一个解,由 $AGA=A$ 则有

$$AGAx = Ax = b,$$

即 $AGb=b$ 成立。这表明 $\tilde{x}=Gb$ 是方程 $Ax=b$ 的解。故由定义 1,G 是 A 的一个$\{1\}$-广义逆。

必要性。设对任意的 $m\times 1$ 矩阵 b,方程 $Ax=b$ 有解 $x=Gb$。对任意 $z\in C^n$,$Az=b$ 是 $m\times 1$ 矩阵。由上述假设得 $AGb=b$,于是

$$AGAz = Az \quad (\forall z \in C^n),$$

因此,$AGA=A$。证毕。

现在的问题是:对给定的 $m\times n$ 矩阵 A,它是否一定有$\{1\}$-广

---

* 本节及下节所讨论的矩阵均指复数矩阵。

义逆 G? 如果有,又如何确定 A 的所有{1}-广义逆?

定理 2 对 $m\times n$ 矩阵 A,设

$$PAQ=\begin{pmatrix} E_r & \\ & 0 \end{pmatrix} \qquad (1)$$

这里 P,Q 各为 m 阶,n 阶可逆矩阵,则 A 的所有{1}-广义逆的集合为

$$A\{1\}=\left\{Q\begin{pmatrix} E_r & X \\ Y & Z \end{pmatrix}P;\text{这里 } X,Y,Z \text{ 分别是任意的 } r\times(m-r),(n-r)\times r,(n-r)\times(m-r) \text{ 矩阵}\right\}.$$

证明 设 A 的秩为 r,则必有 m,n 阶可逆矩阵 P,Q,使得(1)式成立。因而

$$AGA=A\Leftrightarrow PAGAQ=PAQ$$

$$\Leftrightarrow PAQ\cdot Q^{-1}GP^{-1}\cdot PAQ=PAQ$$

$$\Leftrightarrow\begin{bmatrix} E_r & \\ & 0 \end{bmatrix}\cdot\begin{pmatrix} W & X \\ Y & Z \end{pmatrix}\cdot\begin{bmatrix} E_r & \\ & 0 \end{bmatrix}=\begin{bmatrix} E_r & \\ & 0 \end{bmatrix}$$

$$\Leftrightarrow\begin{bmatrix} W & \\ & 0 \end{bmatrix}=\begin{bmatrix} E_r & \\ & 0 \end{bmatrix}$$

$$\Leftrightarrow W=E_r$$

这里记

$$Q^{-1}GP^{-1}=\begin{pmatrix} W & X \\ Y & Z \end{pmatrix},$$

W 是 $r\times r$ 子块。因此,A 的每个{1}-广义逆 G 均具有如下形式

$$G=Q\begin{pmatrix} E_r & X \\ Y & Z \end{pmatrix}P$$

X,Y,Z 如前所述。证毕。

由定理 2 可得知,对任意 $m\times n$ 矩阵 A,它的{1}-广义逆总是

存在，即集合 $A\{1\}$ 非空。又由 X,Y,Z 的任意性，可见一般情况下，A 的 $\{1\}$-广义逆不是唯一的。集合 $A\{1\}$ 中任一确定的元素(即 A 的一个 $\{1\}$-广义逆，上面曾用 G 表示)常记为 A^-。由定义，$AA^-A=A$。

当 $m=n=r$，则 A 可逆，于是有

$$\begin{aligned} A^{-1} = A^{-1}AA^{-1} &= A^{-1}(AA^- A)A^{-1} \\ &= (A^{-1}A)A^- (AA^{-1}) \\ &= A^-. \end{aligned}$$

这对任何 $A^-\in A\{1\}$ 都成立，所以这时 $A\{1\}$ 只有唯一的元素 A^{-1}。由此可见，当 A 可逆时，A 的 $\{1\}$-广义逆 A^- 也就是 A 的通常的逆阵 A^{-1}。

定理 3 若 A^- 是 $m\times n$ 矩阵 A 的一个 $\{1\}$-广义逆，则当方程组 $Ax=b$ 有解时，其通解可表示为

$$x = A^- b + (E_n - A^- A)z,$$

这里 z 是任意的 n 维列向量。

证明 先看齐次方程组 $Ax=\theta$，这时由等式 $A-AA^-A=0$(矩阵)，推得

$$A(E_n - A^- A)z = \theta$$

对任一 $z\in C^n$ 成立。即对任一 $z\in C^n$，$(E_n-A^-A)z$ 一定是 $Ax=\theta$ 的解。要证明 $Ax=\theta$ 的解均可以表示成这种形式，只须证明 E_n-A^-A 的秩为 $n-r$，这里 r 是 A 的秩。由定理 2 可得

$$A^- = Q\begin{pmatrix} E_r & X \\ Y & Z \end{pmatrix}P,$$

而

$$E_n - A^- A = E_n - Q\begin{pmatrix} E_r & X \\ Y & Z \end{pmatrix}PA,$$

故

$$\operatorname{rank}(E_n - A^- A) = \operatorname{rank}(Q^{-1}(E_n - A^- A)Q)$$

$$= \operatorname{rank}\left[E_n - \begin{bmatrix} E_r & X \\ Y & Z \end{bmatrix}\begin{bmatrix} E_r & \\ & 0 \end{bmatrix}\right]$$

$$= \text{rank}\begin{bmatrix} 0 & 0 \\ -Y & E_{n-r} \end{bmatrix} = n - r.$$

对非齐次方程 $Ax=b$,若它有解,则由{1}-广义逆的定义,A^-b 必为解,从而通解是

$$x = A^- b + (E_n - A^- A)z, \quad z \in C^n.$$

定理证毕.

利用上述通解表示式,可以研究相容线性方程组 $Ax=b$ 的解的性质.

对给定的相容线性方程组

$$Ax = b \quad (A \in C^{m\times n},\ b \in C^m),$$

它的解一般不唯一. 而在一些实际问题中需要在它的一切解之中求出使得范数

$$\| x \|_2 = \sqrt{x^H x}$$

为最小的解. 这样的解称为该方程组的**最小范数解**. 下述两定理都是对酉空间 C^n 的向量来说的.

定理 4 设 A^- 是 $m\times n$ 矩阵 A 的一个{1}-广义逆,并且$(A^-A)^H=A^-A$,那末对任给的 m 维列向量 b,只要方程组 $Ax=b$ 有解,则 $x^*=A^-b$ 就是它的最小范数解.

证明 设 $Ax=b$ 有解,则其通解为

$$x = A^- b + (E_n - A^- A)z \quad z \in C^n.$$

现设$(A^-A)^H=A^-A$,今证 $x=A^-b$ 是最小范数解. 由于

$$\begin{aligned}\| x \|_2^2 &= \| A^- b + (E_n - A^- A)z \|_2^2 \\ &= [A^- b + (E_n - A^- A)z]^H[A^- b + (E_n - A^- A)z] \\ &= (A^- b)^H(A^- b) + [(E_n - A^- A)z]^H[(E_n - A^- A)z] \\ &\quad + (A^- b)^H[(E_n - A^- A)z] + [(E_n - A^- A)z]^H(A^- b),\end{aligned}$$

令 $b=Ax_0$,则有

$$(A^- b)^H[(E_n - A^- A)z] = (A^- Ax_0)^H(E_n - A^- A)z$$

$$
\begin{aligned}
&= x_0^H (A^- A)^H (E_n - A^- A) z \\
&= x_0^H (A^- A)(E_n - A^- A) z \\
&= x_0^H (A^- A - A^- A A^- A) z \\
&= x_0^H (A^- A - A^- A) z \\
&= 0.
\end{aligned}
$$

同理可证

$$
[(E_n - A^- A)z]^H (A^- b) = 0.
$$

因此有

$$
\begin{aligned}
\| x \|_2^2 &= (A^- b)^H (A^- b) + [(E_n - A^- A)z]^H [(E_n - A^- A)z] \\
&= \| A^- b \|_2^2 + \| (E_n - A^- A)z \|_2^2 \\
&\geqslant \| A^- b \|_2^2.
\end{aligned}
$$

所以，$A^- b$ 是 $Ax = b$ 的最小范数解。证毕.

上面我们讨论了相容线性方程组 $Ax = b$ 的最小范数解，并由广义逆矩阵 A^- 来描述。然而在不少实际问题中（如数据处理问题，与正态分布有关的统计问题等），所得到的方程组 $Ax = b$ 往往是不相容的，因而不存在求它的解（准确解）的问题，但求它的近似解却有意义。我们可以求 $x \in C^n$，使 $\| Ax - b \|_2$ 为极小。这样的 x 就是 $Ax = b$ 的近似解，称为**最小二乘解**。与最小范数解的情形类似，可以证明下述定理.

定理 5 设 A^- 是 $m \times n$ 矩阵 A 的一个 $\{1\}$-广义逆，并且 $(AA^-)^H = AA^-$，则对任给的 m 维列向量 b，$x^* = A^- b$ 一定是 $Ax = b$ 的最小二乘解.

定理的证明从略。以上我们从相容线性方程组 $Ax = b$ 的解来定义矩阵 A 的 $\{1\}$-广义逆 G 或 A^-。下面给出广义逆的一般定义.

定义 2 设 A 是任一 $m \times n$ 矩阵，如果 $n \times m$ 矩阵 G 满足 Moore—Penrose 方程

(1) $AGA = A$;　　　　(2) $GAG = G$;

(3) $(GA)^H = GA$；　(4) $(AG)^H = AG$

的一部分或全部，则称 G 为 A 的*广义逆矩阵*。

按这定义，广义逆矩阵可以分为满足其中一个方程的广义逆矩阵，满足其中两个或三个、四个方程的广义逆矩阵。这样就共有

$$C_4^1 + C_4^2 + C_4^3 + C_4^4 = 15$$

类广义逆矩阵。

满足条件(1)的广义逆矩阵就是上述的{1}-广义逆，这类广义逆(记为 $A\{1\}$)中任何确定的一个，就记作 A^-。

满足条件(1)及(3)的广义逆矩阵类记为 $A\{1,3\}$，其中任一确定的广义逆矩阵记作 A_m^-。例如，定理 4 中的广义逆 A^- 其实是某个 A_m^-。

满足条件(1)及(4)的广义逆矩阵类记为 $A\{1,4\}$，其中的任一成员记为 A_l^-；如定理 5 中的 A^- 就是某个 A_l^-。

满足全部四个条件的广义逆矩阵类记为 $A\{1,2,3,4\}$。这类广义逆对给定的 A 来说，只含有唯一的一个广义逆，常记为 A^+，并称为 M-P 广义逆。

§5　广义逆矩阵 A^+

本节主要讨论广义逆 A^+ 的存在性、唯一性及运算。由上节定义 2 得知，如果我们证明了 A^+ 的存在性，则其他类型的广义逆的存在性就是显然的了，但是对于它们，唯一性则不成立。

定理 1　设 A 是 $m\times n$ 矩阵，则其 M-P 广义逆矩阵 A^+ 存在而且唯一，即是说，满足下列全部四个条件

(1) $AGA = A$；　(2) $GAG = G$；

(3) $(GA)^H = GA$；　(4) $(AG)^H = AG$

的 $n\times m$ 矩阵 G 存在而且唯一，并记为 A^+。

证明 唯一性。设 G_1, G_2 是两个满足上述四个条件的 $n\times m$ 矩阵，现证 $G_1=G_2$。由于

$$\begin{aligned}G_1 &= G_1AG_1 = (G_1A)G_1 = (G_1A)^HG_1\\&= A^HG_1^HG_1 = (AG_2A)^HG_1^HG_1 = A^HG_2^HA^HG_1^HG_1\\&= (G_2A)^HA^HG_1^HG_1 = G_2AA^HG_1^HG_1\\&= G_2A(G_1A)^HG_1 = G_2AG_1AG_1 = G_2AG_1.\end{aligned}$$

同理可得 $G_2=G_2AG_1$。所以，$G_1=G_2$。

存在性。首先将 A 作奇异值分解（见第三章 § 9）：

$$A = PD^*Q^H,\quad D^* = \begin{bmatrix} D & 0\\ 0 & 0\end{bmatrix}$$

这里 P, Q 是酉矩阵，$D=\mathrm{diag}(d_1, d_2, \cdots, d_r)$，且 $d_1\geqslant d_2\geqslant\cdots\geqslant d_r>0$。

现在证明

$$G = Q\begin{bmatrix} D^{-1} & 0\\ 0 & 0\end{bmatrix}P^H$$

满足定理 1 中的四个条件。事实上，

$$\begin{aligned}AGA &= AQ\begin{bmatrix} D^{-1} & 0\\ 0 & 0\end{bmatrix}P^HA\\&= P\begin{bmatrix} D & 0\\ 0 & 0\end{bmatrix}\begin{bmatrix} D^{-1} & 0\\ 0 & 0\end{bmatrix}\begin{bmatrix} D & 0\\ 0 & 0\end{bmatrix}Q^H\\&= P\begin{bmatrix} D & 0\\ 0 & 0\end{bmatrix}Q^H = A\end{aligned}$$

$$\begin{aligned}GAG &= Q\begin{bmatrix} D^{-1} & 0\\ 0 & 0\end{bmatrix}P^HAQ\begin{bmatrix} D^{-1} & 0\\ 0 & 0\end{bmatrix}P^H\\&= Q\begin{bmatrix} D^{-1} & 0\\ 0 & 0\end{bmatrix}\begin{bmatrix} D & 0\\ 0 & 0\end{bmatrix}\begin{bmatrix} D^{-1} & 0\\ 0 & 0\end{bmatrix}P^H\end{aligned}$$

则

$$
\begin{aligned}
&= Q\begin{bmatrix} D^{-1} & 0 \\ 0 & 0 \end{bmatrix} P^H = G \\
GA &= Q\begin{bmatrix} D^{-1} & 0 \\ 0 & 0 \end{bmatrix} P^H A \\
&= Q\begin{bmatrix} D^{-1} & 0 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} D & 0 \\ 0 & 0 \end{bmatrix} Q^H \\
&= Q\begin{bmatrix} E_r & 0 \\ 0 & 0 \end{bmatrix} Q^H \\
&= \left[Q\begin{bmatrix} E_r & 0 \\ 0 & 0 \end{bmatrix} Q^H\right]^H = (GA)^H \\
AG &= AQ\begin{bmatrix} D^{-1} & 0 \\ 0 & 0 \end{bmatrix} P^H \\
&= P\begin{bmatrix} D & 0 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} D^{-1} & 0 \\ 0 & 0 \end{bmatrix} P^H \\
&= P\begin{bmatrix} E_r & 0 \\ 0 & 0 \end{bmatrix} P^H \\
&= \left[P\begin{bmatrix} E_r & 0 \\ 0 & 0 \end{bmatrix} P^H\right]^H \\
&= (AG)^H.
\end{aligned}
$$

这就证明了 A^+的存在性。定理证毕。

由定理 1 的证明得知,若 $A \in C^{m \times n}$有奇异值分解

$$
A = P\begin{bmatrix} D & 0 \\ 0 & 0 \end{bmatrix} Q^H
$$

则

$$A^+ = Q\begin{bmatrix} D^{-1} & 0 \\ 0 & 0 \end{bmatrix} P^H$$

而且不依赖于 P,Q 的选择，因 A^+ 是唯一的。

若 A 是非奇异方阵(即 A 是可逆方阵)，则可直接验证 A^{-1} 满足定理 1 中的全部条件，故得知 $A^+ = A^{-1}$。所以，A^+ 可以看作是 A^{-1} 的推广。

广义逆矩阵 A^+ 与线性方程组 $Ax=b$ 的最小二乘解也有联系。由于最小二乘解一般是不唯一的，故通常把它们中 2——范数最小的一个称为 $Ax=b$ 的极小最小二乘解(最佳逼近解)。可以证明，$Ax=b$ 必有唯一的极小最小二乘解，而且它就是 $x=A^+b$。

习 题 五

1. 应用§1 定理 1 的两个推论，估计下面矩阵 A 的特征值的界限：

$$A = \begin{bmatrix} 0.9 & 0.01 & 0.12 \\ 0.01 & 0.8 & 0.13 \\ 0.01 & 0.02 & 0.4 \end{bmatrix}$$

2. 利用圆盘定理 1 估计下面矩阵 A 的特征值的分布范围：

$$A = \begin{bmatrix} 1 & -0.5 & -0.5 & 0 \\ -0.5 & 1.5 & i & 0 \\ 0 & -0.5i & 5 & 0.5i \\ -1 & 0 & 0 & 5i \end{bmatrix}$$

3. 证明下面的矩阵 A 的谱半径 $\rho(A) \leqslant 1$：

$$A = \begin{bmatrix} \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{5} & \frac{2}{5} & \frac{1}{5} & \frac{1}{5} \\ \frac{1}{6} & \frac{1}{6} & \frac{3}{6} & \frac{1}{6} \\ \frac{1}{7} & \frac{1}{7} & \frac{1}{7} & \frac{3}{7} \end{bmatrix}$$

4. 在圆盘定理中，如果一个连通部分是由两个外切圆构成的，证明每个圆上不可能都有两个特征值。

5. 设 Q 为酉矩阵，$A=\mathrm{diag}(a_1,a_2,\cdots,a_n)$，证明 QA 的特征值 μ 满足不等式：

$$\min_{1\leqslant i\leqslant n}|a_i|\leqslant|\mu|\leqslant\max_{1\leqslant i\leqslant n}|a_i|.$$

6. 验证 M-P 广义逆矩阵 A^+ 的下列性质：

$$(A^+)^+=A;\quad (AA^+)^2=AA^+;\ (A^H)^+=(A^+)^H.$$

7. 利用公式(设 $A\in C^{m\times n}$)：

$$A^+=\begin{cases}A^H(AA^H)^{-1} & \text{当 } \mathrm{rank}(A)=m,\\ (A^HA)^{-1}A^H & \text{当 } \mathrm{rank}(A)=n,\end{cases}$$

计算下列矩阵的广义逆矩阵 A^+：

$$(1)\quad A=\begin{pmatrix}1 & 2 & 1\\ 0 & 1 & 1\end{pmatrix}\qquad (2)\quad A=\begin{bmatrix}i & 0\\ 1 & i\\ 0 & 1\end{bmatrix}$$

8. 利用公式

$$A^+=\lim_{\varepsilon\to 0}A^H(AA^H+\varepsilon^2E)^{-1}\quad (A\in C^{m\times n})$$

计算矩阵

$$A=\begin{bmatrix}1 & 0 & 0\\ 0 & 1 & -1\\ 1 & 0 & 0\\ 2 & 1 & -1\end{bmatrix}$$

的广义逆 A^+。

9. 证明：线性方程组 $Ax=b$ 有解的充要条件，是 $AA^+b=b$。这里 $A\in C^{m\times n}$，$b\in C^m$。

10. 已知

$$A=\begin{bmatrix}1 & 2\\ 0 & 0\\ 2 & 4\end{bmatrix}\qquad b=\begin{bmatrix}1\\ 0\\ 2\end{bmatrix}$$

求 $Ax=b$ 的最小范数解，及极小最小二乘解。

第六章　非　负　矩　阵

元素都是非负实数的矩阵称为**非负矩阵**。这类矩阵在数理经济学、概率论、弹性系统的微振动理论等许多领域里都有重要的应用。随着非负矩阵应用的日益扩展，它的基本特征已认为是矩阵论的经典性内容之一。为此，本章将介绍非负矩阵的一些基本性质，包括著名的配朗-弗罗比尼乌斯定理，以及与非负矩阵有密切联系而又有特别重要应用价值的闵可夫斯基矩阵(M-矩阵)等有关主要结果。

§1　正矩阵

定义1　$m\times n$ 的实矩阵 $A=(a_{ij})$ 称为非负的(记为 $A\geqslant 0$(矩阵))或正的(记为 $A>0$(矩阵))，如果 A 的所有元素都是非负的(所有 $a_{ij}\geqslant 0$)，或正的(所有 $a_{ij}>0$)。

若 A,B 是两个 $m\times n$ 的实矩阵，x 是 n 维实向量，则由定义1便不难理解 $A\geqslant B$ 及 $x>\theta$ 等式子的含义。

在经济学、概率论及组合学中，有许多矩阵都是非负矩阵或正矩阵。这是不难理解的。这类矩阵有一些为一般矩阵所没有的特殊性质，研究它们显然有重要意义。正矩阵中最重要的结果要算是本节介绍的配朗定理了。

一个方阵 P，如果它的每一行和每一列都只有某个元素为1，其余的元素都为零，则矩阵 P 称为一个**置换矩阵**。

显然交换矩阵的两行(或两列)，可由左乘(或右乘)一个适当

的置换矩阵来实现．置换矩阵是可逆的，且当 P 为置换矩阵时，$P^{-1}=P^T$．

定义 2 矩阵 $A=(a_{ij})_{n\times n}$称为可约的，，如果存在 n 阶置换矩阵 P，使得

$$PAP^T=\begin{pmatrix} A_{11} & 0 \\ A_{21} & A_{22} \end{pmatrix}$$

这里 A_{11}是 k 阶方阵（$1\leqslant k\leqslant n-1$），右上角是 $k\times(n-k)$的零矩阵。否则，就称 A 为不可约的．

按此定义，一阶方阵以及正矩阵都是不可约的．

配朗(Perron)在 1907 年建立了正矩阵的影谱（即矩阵的特征值与特征向量）的卓越的性质，这就是下面的定理．本节的任务主要是证明这一定理．正矩阵的其他较次要的性质，将在后面继续研究．

定理 1(Perron) 任一正矩阵 $A=(a_{ij})_{n\times n}$都有正的特征值 r，它是特征方程的单根，而且大于其他特征值的模．这个“极大”特征值对应于矩阵 A 的一个坐标都是正数的特征向量．

这个定理亦可以叙述成：

定理 1′ 设 $A=(a_{ij})_{n\times n}>0$（矩阵），且 $\rho(A)$为其谱半径，则：

(1) $\rho(A)$为 A 的正特征值，其对应的一个特征向量为正向量；

(2) A 的任何其他特征值 λ，都有 $|\lambda|<\rho(A)$；

(3) $\rho(A)$是 A 的单特征根．

证明 首先来证(1)．设 μ 是 A 的按模最大的一个特征值，又 $x=(x_1,x_2,\cdots,x_n)^T\neq\theta$ 是相应的一个特征向量．于是由定义得

$$Ax=\mu x,\ 且\ |\mu|=\rho(A).$$

现取向量 y 如下：

$$y=(|x_1|,|x_2|,\cdots,|x_n|)^T.$$

今证 y 是 A 的以 $\rho(A)$为特征值的正特征向量．

由于 $Ax=\mu x$．所以对于 $1\leqslant i\leqslant n$ 的整数 i，我们有

$$\mu x_i = \sum_{j=1}^{n} a_{ij} x_j .$$

从而

$$\rho(A)|x_i| = |\mu x_i| \leqslant \sum_{j=1}^{n} a_{ij}|x_j|,$$

写成矩阵形式就是

$$\rho(A)y \leqslant Ay,$$

亦即

$$(A - \rho(A)E)y \geqslant \theta .$$

下面证明这不等式的等号成立。用反证法，设$(A-\rho(A)E)y=z\neq\theta$，则由于 A 是正矩阵，且 z 是非负的非零向量，所以 $Az>\theta$。又显然有 $Ay>\theta$。因此，存在 $\varepsilon>0$，使得 $Az\geqslant\varepsilon Ay$。又由于

$$Az = A(A - p(A)E)y,$$

所以

$$A^2y = Az + \rho(A)Ay \geqslant [\varepsilon + \rho(A)]Ay .$$

令$[\varepsilon+\rho(A)]^{-1}A=B$，则有

$$BAy \geqslant Ay .$$

由于 $B>0$(矩阵)，故由上式可逐步推得

$$B^nAy \geqslant Ay \quad (n = 1,2,\cdots,) \qquad (*)$$

但 B 的谱半径为

$$\rho(B) = \frac{\rho(A)}{\varepsilon + \rho(A)} < 1,$$

所以当 $n\to\infty$时，$B^n\to0$(矩阵)，因而对不等式(*)两边取极限即得 $Ay\leqslant\theta$。这与 $Ay>\theta$ 相矛盾，因此一定有 $z=\theta$。于是我们已证明了

$$Ay = \rho(A)y.$$

这表明 $\rho(A)=|\mu|$是 A 的特征值，而由 y 的选择已知 $y>\theta$，所以 y 是 A 的正特征向量，从而由上式又可得知 $\rho(A)>0$。

现在来证明(2)。为此只须证明除 $\rho(A)$外，不可能有 A 的其他特征值 λ 满足$|\lambda|=\rho(A)$。

若 λ 是 A 的满足$|\lambda|=\rho(A)$的特征值，其相应的特征向量为 u

$=(u_1,u_2,\cdots,u_n)^T$，即 $A\boldsymbol{u}=\lambda\boldsymbol{u}$。又取

$$\boldsymbol{v}=(|u_1|,|u_2|,\cdots,|u_n|)^T,$$

类似(1)的证明过程可得

$$A\boldsymbol{v}=\rho(A)\boldsymbol{v}\ .$$

而由 $A\boldsymbol{u}=\lambda\boldsymbol{u}$ 可得

$$\lambda u_i=\sum_{j=1}^{n}a_{ij}u_j\quad(i=1,2,\cdots,n).$$

两边取绝对值便得

$$\rho(A)|u_i|=|\sum_{j=1}^{n}a_{ij}u_j|\leqslant\sum_{j=1}^{n}a_{ij}|u_j|=(Av)_i.$$

由列向量等式 $A\boldsymbol{v}=\rho(A)\boldsymbol{v}$ 的第 i 个分量相等，又可得

$$(Av)_i=\rho(A)|u_i|\ .$$

于是得到等式

$$|\sum_{j=1}^{n}a_{ij}u_j|=\sum_{j=1}^{n}a_{ij}|u_j|\quad i=1,2,\cdots,n\ .$$

因 $a_{ij}>0$，故上式表明各个 u_j 有相同的幅角 φ，即

$$u_j=|u_j|e^{i\varphi}\quad(i=\sqrt{-1}),j=1,2,\cdots,n$$

而 φ 是不依赖于 j 的常数，因而向量 $\boldsymbol{u}=e^{i\varphi}\boldsymbol{v}$。这表明 $\boldsymbol{u},\boldsymbol{v}$ 只差一个非零常数因子，故 $\boldsymbol{u}$ 也是 A 的以 $\rho(A)$ 为特征值的特征向量(由已证得的等式 $A\boldsymbol{v}=\rho(A)\boldsymbol{v}$ 已得知 $\boldsymbol{v}$ 是 A 的以 $\rho(A)$ 为特征值的特征向量)，但开始时已假设 $\boldsymbol{u}$ 又是 A 的以 λ 为特征值的特征向量，而不同的特征值不能有相同的特征向量，故只能是 $\lambda=\rho(A)$。即(2)得证。

最后来证明(3)。令 $B=\rho^{-1}(A)A=(b_{ij})>0$(矩阵)，则有 $\rho(B)=1$，故要证明(3)，只须证明 1 是 B 的单特征根就行了。这又只须证明在 B 的约当标准形中，相应于特征值 1 的约当块只有一块，并且是一阶子块。

根据结论(1)，必有向量 $\boldsymbol{y}=(y_1,\cdots,y_n)^T>\theta$，使得

$$B\boldsymbol{y}=\boldsymbol{y},\tag{1}$$

从而对任何 $k>1$ 都有

$$B^k y = y . \tag{2}$$

令

$$y_s = \max_i y_i > 0, \quad y_t = \min_i y_i > 0,$$

则由(2)式得

$$y_s \geqslant y_i = \sum_{l=1}^{n} b_{il}^{(k)} y_l \geqslant b_{ij}^{(k)} y_j \geqslant b_{ij}^{(k)} y_t,$$

这里 $b_{ij}^{(k)}$ 代表 B^k 的 (i,j) 元素。从而有

$$b_{ij}^{(k)} \leqslant \frac{y_s}{y_t} .$$

这表明对所有 $k>1$, $b_{ij}^{(k)}$ 是有界的。假若 B 的约当标准形中有一个对应于特征值 1 的约当块的阶数大于 1,不妨设其为 2,则存在可逆矩阵 P,使得

$$B^k = P\begin{bmatrix} 1 & & & & \\ k & 1 & & & \\ & & J_1^k(\lambda_1) & & \\ & & & \ddots & \\ & & & & J_m^k(\lambda_m) \end{bmatrix}P^{-1}$$

它对任何 $k>1$ 成立。这与 $b_{ij}^{(k)}$ 有界相矛盾,故 B 的约当标准形中对应于特征值 1 的约当块是一阶的。下证这种子块只有一个。

设 B 的约当标准形为:

$$J = \begin{bmatrix} E_r & & & \\ & J_1(\lambda_1) & & \\ & & \ddots & \\ & & & J_l(\lambda_l) \end{bmatrix}$$

其中 E_r 为 r 阶单位矩阵,且 $|\lambda_i|<1 \ (i=1,\cdots,l)$。

如果 $r>1$,则令 $C=J-E$。对任一 n 维向量 $x\in C^n$(n 维酉空间),则满足 $Cx=\theta$ 的所有向量 x 的集合形成 C^n 的一个子空间,叫

做 C 的零化子空间。由齐次线性方程组的理论，易知这子空间的维数是 $n-r_c=r$(r_c 是矩阵 C 的秩)。由于 B 与 J 相似，故 r 也是 $B-E$ 的零化子空间的维数。考虑到 $r>1$，故除向量 y 满足上面的(1)式外，必然还有另一向量 z 满足

$$Bz = z,\ z = (z_1, z_2, \cdots, z_n)^T, \qquad (3)$$

并且 z 与 y 线性无关。令

$$\tau = \max_i\left(\frac{z_i}{y_i}\right) = \frac{z_j}{y_j}, \qquad (4)$$

则有 $\tau y \geqslant z$，且不可能取等号，故有

$$B(\tau y - z) > \theta\ .$$

利用(1)式与(3)式，上式可写成

$$\tau y - z > \theta\ .$$

写出上式的第 j 个分量，则有

$$\tau > \frac{z_j}{y_j}\ .$$

这与(4)式定义的 τ 值相矛盾，故 $r=1$。证毕。

§2 非负矩阵

在上节，我们证明了正矩阵的配朗定理，而正矩阵是不可约非负矩阵的一种特殊情形。弗罗比尼乌斯(Frobenius)把上述定理推广到不可约非负矩阵上(1912 年)。本节除介绍这一重要结果外，还讨论非负矩阵的其他一些基本性质。

定理 1(Frobenius) 设 A 是不可约非负矩阵。则 A 总有正的特征值 r，它是特征方程的单根；所有其他特征值的模都不超过 r；这个"极大"特征值 r 对应于矩阵 A 的一个坐标都是正数的特征向量；如果 A 有 h 个特征值

$$\lambda_0 = r, \lambda_1, \cdots, \lambda_{h-1}$$

的模都等于 r，则这些数都是互不相同的而且是方程 $\lambda^h - r^h = 0$（$h \geqslant 1$）的根；复数平面上的点集 $\{r, \lambda_1, \cdots, \lambda_{h-1}\}$ 在绕原点旋转 $\frac{2\pi}{h}$ 角的变换下不变。当 $h>1$ 时，矩阵 A 置换相似于矩阵 D，即

$$PAP^T = D = \begin{bmatrix} 0 & A_{12} & 0 & \cdots & 0 \\ 0 & 0 & A_{23} & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & A_{h-1,h} \\ A_{h1} & 0 & 0 & \cdots & 0 \end{bmatrix}$$

这里 P 是置换矩阵，又 D 的主对角线上的都是非空零方阵。矩阵 D 称为不可约非负矩阵 A 的标准形。

这定理的证明亦非常复杂，这里就不介绍了。

首先要注意的是，定理 1 不能照搬到非负可约矩阵上。但是，由于任一非负矩阵 $A \geqslant 0$（矩阵）都可表示成不可约的正矩阵序列 $\{A_m\}$ 的极限：

$$A = \lim_{m \to \infty} A_m \quad (\text{每个 } A_m > 0(\text{矩阵})), \tag{1}$$

所以不可约非负矩阵的某些性质，在较弱的形式下，对于可约非负矩阵亦能成立。

对于任意的非负矩阵，有下述定理。

定理 2 **设 $A=(a_{ij})_{n \times n}$ 为非负矩阵，则 A 必有一非负特征值 r，而 A 的所有特征值的模都不超过 r；特征值 r 对应于非负特征向量 y：**

$$Ay = ry \quad (y \geqslant \theta, \text{且 } y \neq \theta) .$$

证明 设 A 有表示式(1)，以 $r^{(m)}$ 及 $y^{(m)}$ 各记正矩阵 A_m 的"极大"特征值与相应的单位正特征向量，于是

$$A_m y^{(m)} = r^{(m)} y^{(m)}, \quad (\text{每个 } y^{(m)} > \theta). \tag{2}$$

此时由(1)式知存在极限：

$$\lim r^{(m)} = r .$$

这里 r 为 A 的特征值。因 $r^{(m)}>0$，且 $r^{(m)}>|\lambda_0^{(m)}|$，其中 $\lambda_0^{(m)}$ 为 A_m 的任一特征值($m=1,2,\cdots$)，取极限得：

$$r \geqslant 0, \quad r \geqslant |\lambda_0|,$$

这里 λ_0 为 A 的任一特征值.

又从单位特征向量序列$\{y^{(m)}\}$中可以选出子序列$\{y^{(m_\nu)}\}$，且它收敛于某一单位向量 y. 在等式(2)的两边对 m 的子序列$\{m_\nu\}$取极限，即得

$$Ay = ry \quad (y \geqslant \theta, \text{又 } y \neq \theta) .$$

定理证毕. 注意定理中的 $r=\rho(A)$.

□注：一般的非负矩阵尚有一个比较重要的性质，因证明较复杂，现只把结果列出如下：

若 0(矩阵)$\leqslant B \leqslant A$，则有 $\rho(B) \leqslant \rho(A)$. □

由定义来判别矩阵是否可约是困难的. 有多种判别条件可以使用，下面我们来证明其中一种充要条件，即下述定理.

定理 3 n 阶非负矩阵 A 为不可约的充要条件，是存在正整数 $s \leqslant n-1$，使得

$$(E+A)^s > 0(\text{矩阵}).$$

证明 必要性. 这只须证明对任意向量 $y \geqslant \theta$ $(y \neq \theta)$都有不等式

$$(E+A)^{n-1} y > \theta$$

就行. 我们首先证明在条件 $y \geqslant \theta$ 与 $y \neq \theta$ 下，向量 $z=(E+A)y$ 中零坐标的个数小于向量 y 中零坐标的个数. 假若相反，那末 y 与 z 有相同的零坐标个数(因为 z 的零坐标个数是不会多于 y 的零坐标个数的). 所以，不失一般性，设

$$y = \begin{pmatrix} u \\ 0 \end{pmatrix}, \quad z = \begin{pmatrix} v \\ 0 \end{pmatrix} \quad (u, v > \theta) .$$

这里列向量 u, v 有相同维数. 又令

$$A=\begin{pmatrix}A_{11} & A_{12}\\ A_{21} & A_{22}\end{pmatrix},$$

则有

$$\begin{pmatrix}u\\ 0\end{pmatrix}+\begin{pmatrix}A_{11} & A_{12}\\ A_{21} & A_{22}\end{pmatrix}\begin{pmatrix}u\\ 0\end{pmatrix}=\begin{pmatrix}v\\ 0\end{pmatrix},$$

因此得 $A_{21}u=\theta$. 又因 $u>\theta$,故有 $A_{21}=0$(矩阵)。这与 A 为不可约矩阵相矛盾。所以,z 与 y 有相同的零坐标个数是不可能的,从而证明了向量 z 的零坐标个数小于向量 y 的零坐标个数。

上述结果表明:向量 y $(\theta\leqslant y\neq\theta)$ 每用 $E+A$ 左乘一次,其零坐标个数至少减少一个,因此得

$$(E+A)^{n-1}y>\theta。$$

充分性。设有 $(E+A)^{s}>0$(矩阵),如果 A 是可约的,则存在置换矩阵 P,使得

$$P(E+A)P^{T}=\begin{bmatrix}A_{11}+E_{1} & 0\\ A_{21} & A_{22}+E_{2}\end{bmatrix}$$

$$=\begin{bmatrix}\tilde{A}_{11} & 0\\ A_{21} & \tilde{A}_{22}\end{bmatrix}$$

对任意正整数 k,都有

$$P(E+A)^{k}P^{T}=\begin{bmatrix}\tilde{A} & 0\\ A_{21} & \tilde{A}_{22}\end{bmatrix}^{k}$$

故对所有的上述 k 值,

$$(E+A)^{k}=P^{T}\begin{bmatrix}\tilde{A}_{11} & 0\\ A_{21} & \tilde{A}_{22}\end{bmatrix}^{k}P$$

这等式表明无论正整数 k 为何值,$(E+A)^{k}$ 中永远有零元素,因此 $(E+A)^{s}>0$(矩阵)是不可能的。这就证明 A 不可能是可约的。证毕。

例如,非负矩阵

$$A=\begin{pmatrix}1&1&0\\1&1&1\\0&1&1\end{pmatrix}$$

是不可约的，因为当 $s=3-1=2$ 时，即有

$$(E+A)^2=\begin{pmatrix}2&1&0\\1&2&1\\0&1&2\end{pmatrix}^2=\begin{pmatrix}5&4&1\\4&5&4\\1&4&5\end{pmatrix}>0(\text{矩阵})。$$

我们已经知道正矩阵是不可约的，且是非负矩阵。对于它有下面的结果。

定理 4 设 A 为正矩阵，x 是 A 的对应于特征值 $\rho(A)$ 的特征向量，又 y 是 A^T 的对应于特征值 $\rho(A)$ 的任一正特征向量，则有：

$$\lim_{m\to\infty}(\rho(A)^{-1}A)^m=(y^Tx)^{-1}xy^T。$$

这定理在数理经济学中有直接的应用。（证略）。

现转到非负矩阵的分类问题上。若非负矩阵 A 有 h 个特征值的模均等于 $\rho(A)$，则当 $h=1$ 时，就称方阵 A 是*本原的*；当 $h>1$ 时，就称 A 是*非本原的*。h 统称为 A 的*非本原性指标*。本原矩阵与非本原矩阵的性质很不相同。正矩阵都是本原矩阵，但反之不真。

定理 5 非负矩阵 A 是本原矩阵的充要条件，是存在某个正整数 m，使得 $A^m>0$（矩阵）。（证略）。

例如，非负矩阵

$$A=\begin{pmatrix}0&2\\1&1\end{pmatrix}$$

是本原的，因为

$$A^2=\begin{pmatrix}2&2\\1&3\end{pmatrix}>0(\text{矩阵})。$$

可以证明，对于本原（非负）矩阵 A，配朗定理的结论仍然成立。

推论 设 $A\geqslant0$（矩阵），$B\geqslant0$（矩阵），且 A 为本原矩阵，则：

(1) A^T 也是本原矩阵；

(2) 对任一正整数 p, A^p 也是本原矩阵；

(3) $A+B$ 也是本原矩阵.

证明 (1)与(2)容易证得. 现证(3)如下：由定理 5, 存在正整数 m, 使得 $A^m>0$(矩阵). 又因为

$$(A+B)^m = A^m + C,$$

其中 $C\geqslant 0$(矩阵). 故$(A+B)^m>0$(矩阵). 从而(3)得证.

关于本原矩阵的其他性质及应用，这里就不介绍了.

§3 随机矩阵

这里对非负矩阵中一类很重要的矩阵——随机矩阵作一个简单介绍.

考虑 n 个随机事件组 $S_1,S_2,\cdots,S_n$ 及时间序列 $t_0,t_1,t_2,\cdots$. 若在这些时刻的每一瞬间，这事件组有一个且只有一能够出现. 如果在时刻 t_{k-1} 出现的事件为 S_i，则事件 S_j 在时刻 t_k 出现的概率记为 p_{ij} $(i,j=1,2,\cdots,n;k=1,2,\cdots)$. 又假设条件概率 $p_{ij}(i,j=1,2,\cdots,n)$ 与足数 k 无关.

当给出了条件概率矩阵 $P=(p_{ij})_{n\times n}$时，我们就说给出了有限事件的纯马尔可夫(Markov)链. 在这里显然有

$$p_{ij}\geqslant 0,\ \sum_{j=1}^{n}p_{ij}=1\quad (i,j=1,2,\cdots,n).$$

定义 1 非负矩阵 $A=(a_{ij})_{n\times n}$称为一个随机矩阵，如果 A 的每一行上的元素之和都等于 1.

随机矩阵有重要的应用价值，它在有限马氏过程理论中起着基本的作用，它也经常出现在数理经济学及运筹学的各种模型问题中.

随机矩阵作为非负矩阵中的一类，当然前述非负矩阵的各种

概念和结果，对它也是适用的。这里只是着重考虑它的一些特殊的性质。

从定义得知，随机矩阵 A 有特征值 1，且与之对应的有正特征向量 $z=(1,1,\cdots,1)^T$。反之亦易看到，每个非负的 n 阶矩阵 A 有特征值 1 且对应于 1 的特征向量为 $(1,1,\cdots,1)^T$ 时，则 A 都是随机矩阵。这样我们就得到下述定理。

定理 1 非负矩阵 $A=(a_{ij})_{n\times n}$ 是随机矩阵的充要条件，是矩阵 A 有特征值 1，且 n 维向量 $z=(1,1,\cdots,1)^T$ 是与 1 相应的一个正的特征向量。

可以证明，特征值 1 是随机矩阵的"极大"特征值(其他特征值的模都不超过 1)。

具有正极大特征值与对应正特征向量的非负矩阵，与随机矩阵之间存在着密切的关系。

定理 2 若非负矩阵 $A=(a_{ij})_{n\times n}$ 有正的极大特征值 $r=\rho(A)>0$，且对应于数 r 有正的特征向量 $z=(z_1,z_2,\cdots,z_n)^T>\theta$，则矩阵 A 能相似于数 r 与某个随机矩阵 P 的乘积：

$$A = Z(rP)Z^{-1}, \quad Z = \mathrm{diag}(z_1,\cdots,z_n)\ 。$$

即是说，$(Z^{-1}AZ)/r$ 是随机矩阵。

证明 因为

$$\sum_{j=1}^{n} a_{ij}z_j = rz_i \quad (i = 1,2,\cdots,n) \tag{1}$$

引入对角矩阵

$$Z = \mathrm{diag}(z_1,z_2,\cdots,z_n)$$

及矩阵

$$P = \frac{1}{r}Z^{-1}AZ,$$

则

$$p_{ij} = \frac{1}{r}z_i^{-1}a_{ij}z_j \geqslant 0 \quad (i,j = 1,2,\cdots,n).$$

而由(1)式可得

$$\sum_{j=1}^{n} p_{ij} = 1 \quad (i = 1, 2, \cdots, n) .$$

即 P 是随机矩阵，且有 $A = Z(rP)Z^{-1}$。证毕。

在实际应用中常要考虑随机矩阵 A 的幂序列 $\{A^m\}$ 的收敛性。由于 A 的极大特征值 $r=1$，且 A 的任一模等于 1 的特征值所对应的约当块都是一阶的（证略），故有如下结果。

定理 3 设 A 为不可约随机矩阵，则极限

$$\lim_{m\to\infty} A^m$$

存在的充要条件是 A 为本原矩阵。

例 某区人口流动问题的研究，导致随机矩阵

$$P = \begin{bmatrix} 0 & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} & 0 \end{bmatrix}$$

的幂 P^m 的极限问题。

由于 $E+P>0$，故 P 不可约。又

$$P^2 = \begin{bmatrix} \frac{1}{2} & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{2} \end{bmatrix} > 0$$

所以 P 是本原矩阵，故由定理 3 知 $\lim P^m$ 存在。又用归纳法可证明

$$P^m = \begin{bmatrix} p_m & p_{m+1} & p_{m+1} \\ p_{m+1} & p_m & p_{m+1} \\ p_{m+1} & p_{m+1} & p_m \end{bmatrix}$$

其中

$$p_m = \frac{1}{3}[1 + \frac{(-1)^m}{2^{m-1}}].$$

因此得：

$$\lim_{m \to \infty} P^m = \frac{1}{3}\begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$$

§4 M-矩阵

在生物学、物理学及数理经济学等领域中，有许多问题可以归结为具有特殊构造的矩阵的问题。在这类矩阵中，具有非正的非对角线元素的实方阵扮演着重要角色。设 $A=(a_{ij})_{n\times n}$，其中当 $i\neq j$ 时都有 $a_{ij}\leqslant 0$，易知这种矩阵 A 都可以表示成 $A=sE-B$，其中 $B\geqslant 0$（矩阵），$s>0$。故这种矩阵与非负矩阵有一定的联系。这里讨论其中重要的一类，称为闵可夫斯基(Minkowski)矩阵，简称 M-矩阵。

定义 1 设 $A=sE-B$ 为 n 阶实矩阵，且 $B\geqslant 0, s>0$。那末，若 $s\geqslant\rho(B)$，则称 A 为 M-矩阵；若 $s>\rho(B)$，则称 A 为非奇异 M-矩阵。

全体 n 阶实方阵的集合用记号 $M_n(R)$ 表示。又记

$$Z^{n\times n} = \{A = (a_{ij}) \in M_n(R): 当\ i \neq j, a_{ij} \leqslant 0; i,j = 1,2,\cdots,n\}.$$

我们首先给出非奇异 M-矩阵的一些特性。

定理 1 设 $A\in Z^{n\times n}$ 为非奇异 M-矩阵，且 $D\in Z^{n\times n}$，又 $A\leqslant D$。则：

(1) A^{-1} 与 D^{-1} 存在，且 $A^{-1}\geqslant D^{-1}\geqslant 0$（矩阵）；

(2) D 的每个实特征值为正数；

(3) $|D|\geqslant|A|>0$。

证明 由假设有

$$A = sE - B, \quad B \geqslant 0（矩阵）, s > \rho(B).$$

对任意给定的实数 $\omega\leqslant 0$，考虑矩阵

$$C = A - \omega E = (s-\omega)E - B.$$

由于 $s-\omega>\rho(B)$，故 C 也是非奇异 M-矩阵．这表明非奇异 M-矩阵的每个实特征必是正数．由于 $D\in Z^{n\times n}$，故存在足够小的正数 ε，使得

$$P = E - \varepsilon D \geqslant 0(\text{矩阵}).$$

此时，由条件 $A\leqslant D$，可得

$$Q = E - \varepsilon A \geqslant E - \varepsilon D \geqslant 0(\text{矩阵}),$$

即

$$0(\text{矩阵}) \leqslant P \leqslant Q.$$

因非负矩阵 Q 的谱半径 $\rho(Q)$ 为 Q 的非负特征值，所以

$$|(1-\rho(Q))E - \varepsilon A| = |Q - \rho(Q)E| = 0.$$

由此又得知

$$\frac{1}{\varepsilon}(1-\rho(Q))$$

为 A 的实特征值．由上面所得到的结果(非奇异 M-矩阵的每个特征必是正数)便知

$$1-\rho(Q)>0,$$

因此有

$$0\leqslant \rho(Q)<1.$$

又

$$(\varepsilon A)^{-1} = (E-Q)^{-1} = E + Q + Q^2 + \cdots \geqslant 0(\text{矩阵}).$$

故有 $A^{-1}\geqslant 0$(矩阵)。又因

$$0(\text{矩阵}) \leqslant P^k \leqslant Q^k \quad (k=1,2,\cdots)$$

又

$$\rho(P)\leqslant \rho(Q)<1,$$

于是有

$$(\varepsilon D)^{-1} = (E-P)^{-1} = E + P + P^2 + \cdots \leqslant (\varepsilon A)^{-1}.$$

从而得

$$A^{-1} \geqslant D^{-1} \geqslant 0(\text{矩阵}).$$

即性质(1)已得证.

现证(2):取 $a \leqslant 0$,则 $D-aE \geqslant A$,由(1)得 $D-aE$ 非奇异,因而 D 的所有实特征值为正数。即(2)得证.

最后来证明(3)。由上面的分析,只须证明:若 A 的所有实特征为正数,且 $A \leqslant D$,则(3)便成立。这可以对矩阵的阶数 n 作归纳证明。设 A_1, D_1 分别是 A 及 D 的前 $n-1$ 行、前 $n-1$ 列构成的矩阵.则 A_1, D_1 都属于 $Z^{(n-1)\times(n-1)}$,且 $A_1 \leqslant D_1$.

由于矩阵

$$\tilde{A} = \begin{bmatrix} A_1 & 0 \\ 0 & a_{nn} \end{bmatrix} \in Z^{n\times n}$$

满足 $A \leqslant \tilde{A}$,故由(2),$\tilde{A}$ 的所有实特征值为正数。因而 A_1 的所有实特征值亦为正数。按归纳法假设,即有 $|D_1| \geqslant |A_1| > 0$,而由(1),$A^{-1} \geqslant D^{-1} \geqslant 0$(矩阵)。于是

$$(A^{-1})_{nn} \geqslant (D^{-1})_{nn} \geqslant 0.$$

(这里 $(A^{-1})_{nn}$ 表示 A^{-1} 的 (n,n) 元素.)

亦即

$$\frac{|A_1|}{|A|} \geqslant \frac{|D_1|}{|D|}$$

因此,$|A| > 0$, $|D| > 0$,且有

$$|D| \geqslant |D_1| \cdot |A| / |A_1| \geqslant |A| > 0.$$

定理证毕。

从定理的证明中看到,若 $A, D \in Z^{n\times n}$,且 $A \leqslant D$,则 A 为非奇异 M-矩阵蕴含着 D 也是非奇异 M-矩阵,且有 $|D| \geqslant |A| > 0$。这个结果在许多实际问题中十分有用。此外,若定理 1 中的假设“非奇异 M-矩阵”改换成“A 的每个实特征值都是正数”,则定理 1 的各个结论仍成立。

定义 2 设 $A=(a_{ij})$是 n 阶复数矩阵，如果

$$|a_{ii}| \geqslant \sum_{j\neq i}|a_{ij}| \quad i=1,2,\cdots,n,$$

则称 A 为行对角占优的，如果上述不等式是严格不等的，即

$$|a_{ii}| > \sum_{j\neq i}|a_{ij}| \quad i=1,2,\cdots,n,$$

则称 A 为行严格对角占优的。类似地可以定义列对角占优概念。以下我们把行对角占优简单地说成对角占优。

下面的定理是非奇异 M-矩阵的一个基本定理，它提供了非奇异 M-矩阵的多种等价条件。

定理 2 设 $A\in Z^{n\times n}$，则以下各命题彼此等价：

(1) 存在 $x\geqslant\theta$，使得 $Ax>\theta$；

(2) 存在 $x>\theta$，使得 $Ax>\theta$；

(3) 存在正对角矩阵 D，使得 AD 为严格对角占优矩阵，且 AD 的所有对角线元素为正数；

(4) 若 $B\in Z^{n\times n}$ 且 $B\geqslant A$，则 B 非奇异；

(5) A 的任意主子阵的每个实特征值为正数；

(6) A 的所有主子式为正数；

(7) 对每个 k $(1\leqslant k\leqslant n)$，$A$ 的所有 k 阶主子式之和为正数；

(8) A 的每个实特征值为正数；

(9) A 为非奇异 M-矩阵；

(10) 存在 A 的一种分裂 $A=P-Q$，使得

$P^{-1}\geqslant 0$(矩阵)，$Q\geqslant 0$(矩阵) 且 $\rho(P^{-1}Q)<1$；

(11) A 非奇异且 $A^{-1}\geqslant 0$(矩阵)。

证明 (1)⇒(2)：

设 $x\geqslant\theta$ 满足 $Ax>\theta$。令 $e=(1,1,\cdots,1)^T\in R^n$；由于 $Ax>\theta$，故有 $\varepsilon>0$，使得

$$Ax+\varepsilon Ae>\theta.$$

这时，$x+\varepsilon e>\theta$ 满足 $A(x+\varepsilon e)>\theta$。

(2)⇒(3)：

设 $x>\theta$ 满足 $Ax>\theta$。令 $D=\mathrm{diag}(x_1,x_2,\cdots,x_n)$，则有

$$a_{ii}x_i > \sum_{j\neq i}|a_{ij}|x_j \quad i = 1,2,\cdots,n,$$

因此 AD 是严格对角占优矩阵，且所有对角线元素为正数。

(3)⇒(8)：

设 $D=\mathrm{diag}(d_1,d_2,\cdots,d_n)$ 为(3)中的矩阵，且 AD 严格对角占优，其所有对角线元素为正数，则 A 为广义严格对角占优矩阵，即

$$|a_{ii}| > \frac{1}{x_i}\sum_{j\neq i}|a_{ij}|x_j, \text{ 且 } x_i > 0 \quad (i = 1,2,\cdots,n).$$

又所有 $a_{ii}>0$，故由圆盘定理得知 A 的每个实特征值为正数。

(8)⇒(9)：

设 $A=sE-B,s>0,B\geqslant 0$(矩阵)，则 $s-\rho(B)$ 为 A 的实特征值，由(8)知它是正数，即 $s>\rho(B)$，故 A 为非奇异 M-矩阵。

(9)⇒(4)：由定理 1 即得。

(4)⇒(5)：

设 A_k 为 A 的任一 k 阶主子式，λ 为 A_k 的实特征值。以下用反证法证明 $\lambda>0$。

假若 $\lambda\leqslant 0$，则定义矩阵 $B\in Z^{n\times n}$ 如下：

$$b_{ij} = \begin{cases} a_{ii} - \lambda & \text{当 } i = j, \\ a_{ij} & \text{当 } i \neq j \text{ 且 } i,j \text{ 属于 } A_k \text{ 在 } A \text{ 中的行、列序数集}, \\ 0 & \text{其他情形}. \end{cases}$$

由于 $\lambda\leqslant 0$，故有 $B\geqslant A$。由(4)知 B 非奇异，又因 $B_k=A_k-\lambda E_k$，且 λ 为 A_k 的特征值，所以

$$|B| = \prod_{i\in S} b_{ii}|B_k| = 0.$$

这里 B_k 为 B 的 k 阶主子阵，它在 B 中的行、列数与 A_k 相同，而 S 表示这种行数(也是列数)的集合。比如当 $S=\{1,3,7\}$ 时，则 B_3 为 B 的由第 1、3、7 行及第 1、3、7 列的元素构成的 3 阶主子阵。由 $|B|=$

0 知 B 是奇异的，这与(4)中 B 为非奇异的假设相矛盾，故有 $\lambda>0$。

(5)⇒(6)：

因为若 A_k 为 A 的任一 k 阶主子阵，则

$$f(\lambda)=|\lambda E_k-A_k|=\lambda^k+a_1\lambda^{k-1}+\cdots+(-1)^k|A_k|。$$

而 $|A_k|=\lambda_1\lambda_2\cdots\lambda_k$；由于 A_k 是实矩阵，$f(\lambda)$是实系数多项式，故若有复根则必成双。因此，在 $|A_k|=\lambda_1\lambda_2\cdots\lambda_k$ 的乘积中，如有复数的特征值 λ_j，则必有另一复数特征值 $\bar{\lambda}_j$(共轭复数)，而 $\lambda_j\bar{\lambda}_j>0$。再由(5)便知 $|A_k|>0$。

(6)⇒(7)：显然成立。

(7)⇒(8)：

因为

$$|A-\lambda E|=(-\lambda)^n+c_1(-\lambda)^{n-1}+\cdots+c_n, (*)$$

式中 c_k 为 A 的所有 k 阶主子式之和。由(7)知所有 $c_k>0$，因此(*)式不能有非正的实根，亦即 A 的所有实特征值都是正数。

(9)⇒(10)：

取 $P=sE$ 及 $Q=B$，这里的 s，B 满足

$$A=sE-B,\ s>\rho(B),B\geqslant 0(\text{矩阵})。$$

这时，$P^{-1}\geqslant 0$(矩阵)，$Q\geqslant 0$(矩阵)，又

$$\rho(P^{-1}Q)=\rho(\frac{1}{s}B)=\frac{1}{s}\rho(B)<1.$$

(10)⇒(11)：

设(10)成立，则有 $A=P(E-C)$，其中 $C=P^{-1}Q$，又因 $\rho(C)<1$，所以有

$$A^{-1}=(E-C)^{-1}P^{-1}=(E+C+C^2+\cdots)P^{-1},$$

故从 $C=P^{-1}Q\geqslant 0$(矩阵)，得 $A^{-1}\geqslant 0$(矩阵)。

(11)⇒(1)：

令 $x=A^{-1}e$，这里 $e=(1,1,\cdots,1)^T$，由(11)有 $A^{-1}\geqslant 0$(矩阵)，所以 $x\geqslant\theta$，$Ax=e>\theta$。定理证毕。

应用矩阵的不可约性质，常可得到一些加强的结果。对于$A\in Z^{n\times n}$为不可约矩阵的情形，则有

定理 3 设$A\in Z^{n\times n}$且为不可约矩阵，则下列各命题彼此等价：

(1) 存在$x>\theta$，使得$Ax>\theta$；

(2) 存在$x>\theta$，使得$Ax\geqslant\theta$，$Ax\neq\theta$；

(3) A为非奇异M-矩阵；

(4) $A^{-1}>0$(矩阵)。

定理证明从略。

以上讨论了非奇异M-矩阵的一些基本性质，但一般的M-矩阵与非奇异的M-矩阵在应用中几乎同等重要。由于一般的M-矩阵(尤其是奇异的M-矩阵)研究的难度大，故其理论比起非奇异M-矩阵来要弱一些。

下面，我们给出一般M-矩阵的一些特性。

定理 4 设$A\in Z^{n\times n}$，则以下各命题彼此等价：

(1) A是M-矩阵；

(2) 对每个$\varepsilon>0$，$A+\varepsilon E$是非奇异M-矩阵；

(3) A的任意主子阵的每个实特征值非负；

(4) A的所有主子式非负；

(5) 对每个$k=1,2,\cdots,n$，A的所有k阶主子式之和为非负实数；

(6) A的每个实特征值非负。

证明 (1)$\Rightarrow$(2)：由M-矩阵定义即得。

(2)$\Rightarrow$(3)：

设A_k为A的任一k阶主子阵，λ是A_k的特征值。若λ为负实数，则由(2)，$B=A-\lambda E$为非奇异M-矩阵，因而B_k的所有实特征值为正数。但是，0是$B_k=A_k-\lambda E_k$的特征值，故得出矛盾，即(3)成立。

(3)⇒(4)：

因 A 的主子式等于对应主子阵所有特征值的乘积，而非实特征值之积为正数。

(4)⇒(5)：显然成立。

(5)⇒(6)：

类似于定理 2 证明中(7)⇒(8)的证法，并利用此证法中的(*)式，这时(5)保证此公式中的 $c_k \geqslant 0$ $(k=1,2,\cdots,n)$。

如果(6)式不成立，则 A 有实的负特征值 λ_0，将 λ_0 代入(*)式，则左边为零，右边为非负实数之和，且第一项 $(-\lambda_0)^n > 0$，这是不可能的。

(6)⇒(1)：

用类似于定理 2 中证明(8)⇒(9)的方法即得。定理证毕。

类似于定理 3 的情形，我们有：

定理 5　设 A 为不可约的奇异 M－矩阵，则：

(1) $\operatorname{rank} A = n-1$；

(2) 存在正向量 $x > \theta$，使得 $Ax = \theta$；

(3) A 的所有真主子阵为非奇异的 M-矩阵，特别有 $a_{ii} > 0$ $(1 \leqslant i \leqslant n)$；

(4) 对任意 $x \in R^n$，若 $Ax \geqslant \theta$，则 $Ax = \theta$。

证明　这里只给出(1)、(2)、(4)的证明。

(1) 设 $A = sE - B$，$s > 0$，$B \geqslant 0$(矩阵)。则 B 为 n 阶不可约非负矩阵，故由 Frobenius 定理，$\rho(B) > 0$ 为 B 的单特征根，又因 A 是奇异的，故 $s - \rho(B) = 0$ 为 A 的单特征根，于是有 $\operatorname{rank} A = n-1$。

(2) 设 $x > \theta$ 且 $Bx = \rho(B)x$，于是

$$Ax = sx - Bx = \rho(B)x - Bx = \theta.$$

(4) 设对某个 $x \in R^n$，$Ax \geqslant \theta$。由(2)有 $y > \theta$ 满足 $A^T y = \theta$ 或 $y^T A = \theta$。若 $Ax \neq \theta$，则有 $y^T Ax \neq 0$，这与 $y^T A = \theta$ 矛盾。

最后略提一下 M-矩阵在投入-产出分析中的应用。

以上述的M矩阵理论为基础，经过简单的分析推导可得到开式Leontief模型(一种基本的投入-产出生产模型)的基本定理：

若此模型的投入矩阵为非负矩阵 T，且 $A=E-T$，则下列各个命题互相等价：

(1) A 为非奇异M-矩阵；

(2) 模型是可行的；

(3) 模型是可获利的。

对于闭式Leontief模型，要用到"具有性质 c"的M-矩阵的概念。一个M-矩阵 A 称为"具有性质 c"的，如果 A 可以表示为 $A=sE-B, s>0, B\geqslant 0$(矩阵)，且矩阵 $T=\frac{1}{s}B$ 的幂序列收敛。

关于闭式Leontief模型的可行性有下述结果：

(1) 具有投入矩阵 T 的闭式Leontief模型为可行的，其必要条件是矩阵 $A=E-T$ 为"具有性质 c"的M-矩阵；

(2) 当矩阵 T 不可约且 A 奇异，则模型可行的充要条件，是 A 为"具有性质 c"的M-矩阵。

关于M-矩阵的应用，这里就不作更深入的讨论。

参 考 书 目

[1] Ф.Р. 甘特马赫尔.矩阵论,高等教育出版社,1955.

[2] И.М. 盖尔冯德,线性代数学,高等教育出版社,1957.

[3] 北京大学数学系,高等代数,人民教育出版社,1978.

[4] 蒋尔雄等,线性代数,人民教育出版社,1979.

[5] 黄 琳,系统与控制理论中的线性代数,科学出版社.1984.

[6] 陈公宁,矩阵理论与应用,高等教育出版社,1990.

[7] 丁学仁等,工程中的矩阵理论,天津大学出版社,1985.

[8] R.A. 台恩等,矩阵分析,天津大学出版社,1989.

[9] 李 乔,矩阵论八讲,上海科学技术出版社,1988.

[10] 蒋正新等,矩阵理论及其应用,北京航空学院出版社,1988.

[11] R. Bellman, Introduction to Matrix Analysis, McGraw-Hill, N. Y., 1970.

[12] A. S. Deif, Advanoed Matrix Theory for Scientists and Engineers, Abacus Press, 1982.

[13] 郑大钟,线性系统理论,清华大学出版社,1990.

[14] 何 钺,现代控制理论基础,机械工业出版社,1987.

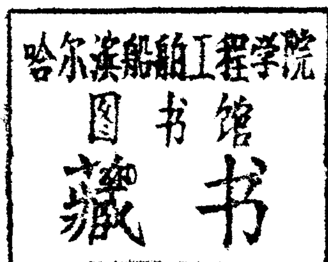
[15] 蔡大用,数值代数,清华大学出版社,1987.

[16] И.С,索明斯基等,高等代数习题集,高等教育出版社,1956.

[17] И.В. 普罗斯库烈柯夫,线性代数习题集,人民教育出版社,1981.

[18] 杨子胥,高等代数习题解,山东科学技术出版社,1982.

[19] C. R. Rao and S. K. Mitra, Generalized Inverse of Matrices and its Applications, JOHN WILEY & SONS, INC, 1971.



$$=a_1\boldsymbol{e}_1+\cdots+a_n\boldsymbol{e}_n=\boldsymbol{x}$$

5.(证)只需证明 $\boldsymbol{e},\cdots,\boldsymbol{e}_n$ 线性无关。设 V 中有向量 $\boldsymbol{x}_0=\boldsymbol{x}_1\boldsymbol{e}_1+\cdots+\boldsymbol{x}_n\boldsymbol{e}_n$ 且表示法唯一。

如有

$$c_1\boldsymbol{e}_1+\cdots+c_n\boldsymbol{e}_n=\theta,$$

则有

$$\boldsymbol{x}_0=(x_1+c_1)\boldsymbol{e}_1+\cdots+(x_n+c_n)\boldsymbol{e}_n,$$

由于向量 $\boldsymbol{x}_0$ 表示法唯一，故有

$$x_1=x_1+c_1,\cdots,x_n=x_n+c_n,$$

从而有 $c_1=0,\cdots,c_n=0.$

这证明了 $\boldsymbol{e}_1,\cdots,\boldsymbol{e}_1$ 线性无关。

6. 1) $\left(1,\frac{1}{2},-\frac{1}{2}\right)$

2) $(33,-82,154)$

7.(1)过渡矩阵为

$$A=\begin{pmatrix}2&0&5&6\\1&3&3&6\\-1&1&2&1\\1&0&1&3\end{pmatrix}$$

(2)向量 $\boldsymbol{x}=(\xi_1,\xi_2,\xi_3,\xi_4)$ 在基 $\beta_1,\beta_2,\beta_3,\beta_4$ 下的坐标为：

$$\begin{cases}\xi_1'=\frac{4}{9}\xi_1+\frac{1}{3}\xi_2-\xi_3-\frac{11}{9}\xi_4,\\ \xi_2'=\frac{1}{27}\xi_1+\frac{4}{9}\xi_2-\frac{1}{3}\xi_3-\frac{23}{27}\xi_4,\\ \xi_3'=\frac{1}{3}\xi_1\qquad\qquad-\frac{2}{3}\xi_4,\\ \xi_4'=-\frac{7}{27}\xi_1-\frac{1}{9}\xi_2+\frac{1}{3}\xi_3+\frac{26}{27}\xi_4.\end{cases}$$

(3)非零向量为 $(c,c,c,-c)$,(任 $c\neq0$)

8.(1)是子空间;(2)不是子空间。

9.(提示)。因为

$$(2,-1,3,3)=(-1)(1,1,0,0)+3(1,0,1,1),$$
$$(0,1,-1,-1)=(1,1,0,0)+(-1)(1,0,1,1).$$

10.(证)如 V_1 的维数是 0,则 V_1 与 V_2 都是零空间,当然相等。如 V_1 的维数是 $m\neq 0$,由于 $V_1\subseteq V_2$ 故 V_1 的任一组基

$$\boldsymbol{e}_1,\boldsymbol{e}_2,\cdots,\boldsymbol{e}_m$$

都是 V_2 中的线性无关组。又因 V_2 与 V_1 的维数相同,故这个线性无关组也是 V_2 的一组基。V_1 与 V_2 有相同的基,因此相等。

11. 设 $\boldsymbol{x}=(a_1,a_2,a_3,a_4)\in V\cap W$,则有

$$a_1-a_2+a_3-a_4=0,\quad a_1+a_2+a_3+a_4=0.$$

由此二式相加或相减便可得:

$$a_1+a_3=0,\quad a_2+a_4=0$$

从而 $a_1=-a_3,a_2=-a_4$,故得

$$\boldsymbol{x}=(a_1,a_2,-a_1,-a_2)$$
$$=a_1(1,0,-1,0)+a_2(0,1,0,-1)$$

但(1,0,−1,0),(0,1,0,−1)线性无关,它就是所求的基。

12. 因 $V_1+V_2=L(\alpha_1,\alpha_2,\alpha_3)+L(\beta_1,\beta_2)$

$$=L(\alpha_1,\alpha_2,\alpha_3,\beta_1,\beta_2).$$

故 V_1+V_2 的维数就是向量组

$$\alpha_1,\alpha_2,\alpha_3,\beta_1,\beta_2$$

的秩,它的任一最大线性无关组,都构成 V_1+V_2 的一组基,以 α_1,$\alpha_2,\alpha_3,\beta_1,\beta_2$ 为行向量构成矩阵 A,并施行初等行变换,便得到:

$$A=\begin{pmatrix}1&0&2&1\\2&0&1&-1\\3&0&3&0\\1&1&0&1\\4&1&3&1\end{pmatrix}\Rightarrow\begin{pmatrix}1&0&2&1\\0&0&-3&-3\\0&0&-3&-3\\0&1&-2&0\\0&1&-5&-3\end{pmatrix}$$

$$\Rightarrow\begin{bmatrix}1&0&2&1\\0&0&1&1\\0&0&0&0\\0&1&-2&0\\0&0&-3&-3\end{bmatrix}\Rightarrow\begin{bmatrix}1&0&2&1\\0&0&1&1\\0&0&0&0\\0&1&-2&0\\0&0&0&0\end{bmatrix}$$

由此可见矩阵 A 的秩等于3，且 $\alpha_1,\alpha_2,\beta_1$ 线性无关，它就是 V_1+V_2 的一组基。

13.(1) 设 $\alpha,\beta\in V_1$，且

$$\alpha=\sum_1^n x_i\boldsymbol{e}_i=\sum_1^n x_i\varepsilon_i$$

$$\beta=\sum_1^n y_i\boldsymbol{e}_i=\sum_1^n y_i\varepsilon_i.$$

则有

$$\alpha+\beta=\sum_1^n(x_i+y_i)\boldsymbol{e}_i=\sum_1^n(x_i+y_i)\varepsilon_i,$$

$$k\alpha=\sum_1^n kx_i\boldsymbol{e}_i=\sum_1^n kx_i\varepsilon_i\quad(k\text{ 是数})$$

即 $\alpha+\beta$ 与 $k\alpha$ 在两组基上的坐标也是相同的，所以 $\alpha+\beta\in V_1$，$k\alpha\in V_1$，即 V_1 是子空间。

(2) 因 V 中每个向量在两组基下的坐标相同，所以基向量 $\boldsymbol{e}_i(i=1,2,\cdots,n)$ 在 $\boldsymbol{e}_1,\boldsymbol{e}_2,\cdots,\boldsymbol{e}_n$ 下的坐标为 $(0,\cdots,0,1,0\cdots,0)$（第 i 个坐标为1）它也应为 $\boldsymbol{e}_i$ 在基 $\varepsilon_1,\varepsilon_2,\cdots,\varepsilon_n$ 下的坐标，于是有：

$$\boldsymbol{e}_i=1\varepsilon_i=\varepsilon_i.\ (i=1,2,\cdots,n)$$

14. 由齐次线性方程组的理论可推知 V_1 是 $n-1$ 维的，且有基 $\alpha_1=(-1,1,0,\cdots,0),\alpha_2=(-1,0,1,0,\cdots,0),\cdots,\alpha_{n-1}=(-1,0,\cdots,0,1)$。又 $x_1=x_2=\cdots=x_n$ 即

$$\begin{cases} x_1 - x_2 & = 0 \\ \quad x_2 - x_3 & = 0 \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots & \\ \qquad x_{n-1} - x_n & = 0 \end{cases}$$

这方程组系数矩阵的秩为 $n-1$，故解空间 V_2 的维数为 1，令 $x_n = 1$，便得 V_2 的一组基，即 n 维向量 $\beta = (1,1,\cdots,1)$，又以 $\alpha_1,\alpha_2,\cdots,\alpha_{n-1},\beta$ 为行的 n 阶行列式

$$\begin{vmatrix} -1 & 1 & 0 & \cdots & 0 & 0 \\ -1 & 0 & 1 & \cdots & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ -1 & 0 & 0 & \cdots & 0 & 1 \\ 1 & 1 & 1 & \cdots & 1 & 1 \end{vmatrix} = (-1)^{n+1} \cdot n \neq 0.$$

故 $\alpha_1,\alpha_2,\cdots,\alpha_{n-1},\beta$ 为 P^n 的一组基，且有 $P^n = V_1 \oplus V_2$。

15. 设 α 是 $Ax = \theta$ 的一个解（向量），则有

$$\begin{pmatrix} A_1 \\ A_2 \\ \vdots \\ A_S \end{pmatrix} \alpha = \theta, \text{即} \begin{pmatrix} A_1\alpha \\ A_2\alpha \\ \vdots \\ A_S\alpha \end{pmatrix} = \theta,$$

从而有 $A_i\alpha = \theta$，即 $\alpha \in V_i (i = 1,2,\cdots,S)$，故有

$$\alpha \in V_1 \cap V_2 \cap \cdots \cap V_S, \qquad (*)$$

反过来，如有关系式(*)，则将上述过程倒推，便可得出 $A\alpha = \theta$，故 $\alpha \in V$。

综合以上结果便得所证。

16. 设 V 是 n 维线性空间，$e_1,e_2,\cdots,e_n$ 为其一组基，则 $L(e_i)$ 都是一维子空间 $(i = 1,2,\cdots,n)$，且有 $L(e_1) + L(e_2) + \cdots + L(e_n) = L(e_1,e_2,\cdots,e_n) = V$。又因 $e_1,e_2,\cdots,e_n$ 是基，零向量 θ 表示式唯一，故这个和是直和，即 $L(e_1) \oplus L(e_2) \oplus \cdots \oplus L(e_n) = V$。

17. 从定义出发可证明 T_1,T_2 是线性变换，又

$$
\begin{aligned}
(T_1+T_2)(x_1,x_2) &= T_1(x_1,x_2)+T_2(x_1,x_2) \\
&= (x_2,-x_1)=(x_1,-x_2) \\
&= (x_1+x_2,-x_1-x_2)
\end{aligned}
$$

$$
\begin{aligned}
(T_1T_2)(x_1,x_2) &= T_1(T_2(x_1,x_2)) \\
&= T_1(x_1,-x_2)=(-x_2,-x_1),
\end{aligned}
$$

$$
\begin{aligned}
(T_2T_1)(x_1,x_2) &= T_2(T_1(x_1,x_2)) \\
&= T_2(x_2,-x_1)=(x_2,x_1)
\end{aligned}
$$

18. 1)因 $T(A+B)=C(A+B)-(A+B)C$

$$
\begin{aligned}
&= (CA-AC)+(CB-BC) \\
&= T(A)+T(B)
\end{aligned}
$$

$$T(kA)=C(kA)-(kA)C=k(CA-AC)=kT(A)$$

故 T 是线性变换。

2)因为

$$
\begin{aligned}
T(A)\cdot B+A\cdot T(B) &= (CA-AC)\cdot B+A\cdot(CB-BC) \\
&= CAB-ABC=T(AB).
\end{aligned}
$$

19. 因为 $(T+S)(1,0,0)=T(1,0,0)+S(1,0,0)$

$$=(1,0,0)+(0,0,1)=(1,0,1)$$

同理得$(T+S)(0,1,0)=(2,0,0)$,$(T+S)(0,0,1)=(1,1,0)$

这表明线性变换 $T+S$ 把 R^3 的一组基向量

$$\boldsymbol{e}_1=(1,0,0),\quad \boldsymbol{e}_2=(0,1,0),\quad \boldsymbol{e}_3=(0,0,1)$$

变换成 R^3 中的另一组基向量:(易证它们线性无关)

$$\varepsilon_1=(1,0,1),\quad \varepsilon_2=(2,0,0),\quad \varepsilon_3=(1,1,0)$$

又象集 $(T+S)(R^3)$ 是 R^3 的子空间(见 §5 例 6),而 $\varepsilon_1,\varepsilon_2,\varepsilon_3$ 是子空间$(T+S)(R^3)$ 的最大线性无关组,故这个子空间的维数是 3,由第 10 题的结果可知$(T+S)(R^3)=R^3$。(在第 10 题中取 $V_2=V$)

20. 因 $T^2(x,y,z)=T(0,x,y)=(0,0,x)$,所以 T^2 的象集(子空间)是一维的,而 $\boldsymbol{e}_3=(0,0,1)$ 是它的基。由 $T^2(x,y,z)=$

$(0,0,x)=(0,0,0)=\theta$，便知 T^2 的核是 R^3 中一切由形如 $(0,y,z)$ 的向量组成的 2 维子空间，又 $\alpha=(0,1,0)$，$\beta=(0,0,1)$ 是它的一组基。

21. 1) $\begin{pmatrix} 2 & -1 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 0 \end{pmatrix}$　　2) $\begin{pmatrix} -1 & 1 & -2 \\ 2 & 2 & 0 \\ 3 & 0 & 2 \end{pmatrix}$

22. 1)，2)，3)的矩阵均为

$$\frac{1}{2}\begin{pmatrix} -4 & -3 & 3 \\ 2 & 3 & 3 \\ 2 & 1 & -5 \end{pmatrix}$$

23. 因为 $BA=A^{-1}(AB)A$，故 $AB\sim BA$.

24. 设 $B=T_1^{-1}AT_1$，$D=T_2^{-1}CT_2$，则有

$$\begin{pmatrix} B & \\ & D \end{pmatrix}=\begin{pmatrix} T_1 & \\ & T_2 \end{pmatrix}^{-1}\cdot\begin{pmatrix} A & \\ & C \end{pmatrix}\cdot\begin{pmatrix} T_1 & \\ & T_2 \end{pmatrix}$$

25. 1)由线性变换定义，易证 T 是线性变换，又因

$$T^2(x_1,x_2,\cdots,x_n)=T(0,x_1,x_2,\cdots,x_{n-1})$$
$$=(0,0,x_1,x_2,\cdots,x_{n-2}),$$
$$\cdots\cdots\cdots\cdots\cdots\cdots$$
$$T^n(x_1,x_2,\cdots,x_n)=(0,0,\cdots,0).$$

即 $T^n=0$(零变换)。

2) 若

$T(x_1,x_2,\cdots,x_n)=(0,x_1,\cdots,x_{n-1})=(0,0,\cdots,0)$，则 $x_1=x_2=\cdots=x_{n-1}=0$. 即 $T^{-1}(\theta)$ 为由一切形如 $(0,\cdots,0,x_n)$ 的向量构成的子空间，它是一维子空间，$(0,\cdots,0,1)$ 是它的基。

又由 §5 定理 1，便得 $T(V)$ 的维数等于 $n-1$。

习　题　二

1. (1)与(2)均不构成欧氏空间；(3)是。

2. 可验证内积定义的四个条件均满足。

3. $\pm\frac{1}{\sqrt{26}}(4,0,1,-3)$提示：向量 $x=(x_1,x_2,x_3,x_4)\neq\theta$ 与所给三个向量正交的充要条件，是 x_1,x_2,x_3,x_4 为方程组

$$\begin{cases}x_1+x_2-x_3+x_4=0\\x_1-x_2-x_3+x_4=0\\2x_1+x_2+x_3+3x_4=0\end{cases}$$

的非零解。可求得一个非零解为 $\alpha=(4,0,1,-3)$.

4. 令

$$\sum_{i=1}^{n}k_i\alpha_i=\theta$$

则对每个 $\alpha_j(j=1,2,\cdots,n)$ 取内积便得

$$0=(\theta,\alpha_j)=(\sum_{1}^{n}k_i\alpha_i,\alpha_j)=\sum_{1}^{n}k_i(\alpha_i,\alpha_j)$$
$$(j=1,2,\cdots,n)$$

这是关于 $k_1,k_2,\cdots,k_n$ 的齐次线性方程组，它只有零解的充要条件是系数行列式 $\neq 0$。

5. 由不等式 $|(x,y)|\leqslant|x|\cdot|y|$，又取

$$x=(|a_1|,|a_2|,\cdots,|a_n|),\quad y=(1,1,\cdots,1)$$

在 R^n 中展开便得。

6. 可验证

$$(\alpha_i,\alpha_j)=\begin{cases}0, & i\neq j\\1, & i=j\end{cases}$$

7. 先求得基础解系(即解空间的一组基)：

$$\alpha_1=(0,1,1,0,0),$$
$$\alpha_2=(-1,1,0,1,0),$$
$$\alpha_3=(4,-5,0,0,1).$$

再将其正交化、单位化，即得所求的标准正交基：

$$\eta_1 = \frac{1}{\sqrt{2}}(0,1,1,0,0),$$

$$\eta_2 = \frac{1}{\sqrt{10}}(-2,1,-1,2,0),$$

$$\eta_3 = \frac{1}{\sqrt{315}}(7,-6,6,13,5).$$

8.(1)由于 $(\theta,\alpha)=0$,故 $\theta \in V_1$,即 V_1 非空,又若 $\alpha_1,\alpha_2 \in V_1$,$k$ 是实数,则有

$$(\alpha_1+\alpha_2,\alpha)=(\alpha_1,\alpha)+(\alpha_2,\alpha)=0+0=0,$$

$$(k\alpha_1,\alpha)=k(\alpha_1,\alpha)=0.$$

故 $\alpha_1+\alpha_2$ 及 $k\alpha_1$ 均为 V_1 的向量。

(2) 因 $\alpha \neq \theta$,故可将其扩充成 V 的正交基:

$$\alpha_1,\beta_1,\beta_2,\cdots,\beta_{n-1}$$

故 $\beta_1,\beta_2,\cdots,\beta_{n-1} \in V_1$,所以 V_1 至少是 $n-1$ 维的,又因为 $\alpha \neq \theta$,故 $(\alpha,\alpha)>0$ 即 α 不属于 V_1;故 $V_1 \neq V$。所以 V_1 是 $n-1$ 维的子空间。

9. 设 $T(\alpha_3)=x_1\alpha_1+x_2\alpha_2+x_3\alpha_3$,则正交变换 T 在标准正交基 $\alpha_1,\alpha_2,\alpha_3$ 下的矩阵是正交矩阵,即在

$$\begin{cases} T(\alpha_1)=\frac{2}{3}\alpha_1+\frac{2}{3}\alpha_2-\frac{1}{3}\alpha_3 \\ T(\alpha_2)=\frac{2}{3}\alpha_1-\frac{1}{3}\alpha_2+\frac{2}{3}\alpha_3 \\ T(\alpha_3)=x_1\alpha_1+x_2\alpha_2+x_3\alpha_3 \end{cases}$$

中,矩阵

$$A=\begin{pmatrix} \frac{2}{3} & \frac{2}{3} & x_1 \\ \frac{2}{3} & -\frac{1}{3} & x_2 \\ -\frac{1}{3} & \frac{2}{3} & x_3 \end{pmatrix}$$

是正交阵，即有 $A \cdot A^T = E$，于是可解得

$$x_1 = \frac{1}{3}, \quad x_2 = x_3 = -\frac{2}{3}$$

从而就确定了正交变换 T。

10. 设 T_1, T_2 是两个正交变换，对任何 $\boldsymbol{x} \in V$，则有

$$|T_1T_2(\boldsymbol{x})| = |T_1(T_2(\boldsymbol{x}))| = |T_2(\boldsymbol{x})| = |(\boldsymbol{x})|,$$

$$|\boldsymbol{x}| = |T_1T_1^{-1}(\boldsymbol{x})| = |T_1^{-1}(\boldsymbol{x})|.$$

故 T_1T_2 及 T_1^{-1} 均为正交变换。

11. 上三角可逆方阵的逆阵也是上三角方阵，于是可设

$$A^{-1} = \begin{pmatrix} b_{11} & b_{12} & \cdots & b_{1n} \\ 0 & b_{22} & \cdots & b_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & b_{nn} \end{pmatrix}$$

由 $A^T = A^{-1}$，得 $a_{ij} = 0 (i < j)$ 故

$$A = \begin{pmatrix} a_{11} & & & \\ & a_{22} & & \\ & & \ddots & \\ & & & a_{nn} \end{pmatrix}$$

又由 $A^TA = E$，得

$$a_{ii}^2 = 1, \quad (i = 1, 2, \cdots, n)$$

从而 $a_{ii} = \pm 1$，(a_{ii} 是实数)。

12. 若 A 为酉矩阵，则

$$(\boldsymbol{x}A)(\boldsymbol{x}A)^H = \boldsymbol{x}(AA^H)\boldsymbol{x}^H = \boldsymbol{x}\boldsymbol{x}^H,$$

即 $|\boldsymbol{x}A|^2 = |\boldsymbol{x}|^2$，故 $|\boldsymbol{x}A| = |\boldsymbol{x}|$

反过来，由 $|\boldsymbol{x}A| = |\boldsymbol{x}|$ 可得

$$(\boldsymbol{x}A)(\boldsymbol{x}A)^H = \boldsymbol{x}\boldsymbol{x}^H,$$

于是 $\boldsymbol{x}(AA^H - E)\boldsymbol{x}^H = 0$. 令 $B = AA^H - E$，于是有

$$f = \boldsymbol{x}B\boldsymbol{x}^H = \sum_{i,j=1}^{n} b_{ij}x_i\bar{x}_j = 0$$

取 x 满足条件

$$x_i = 1, x_j = 1, \text{其他的 } x_k = 0$$

则得 $b_{ij} = 0$。由于 i, j 的任意性，故知 $B = (b_{ij})$ 是零矩阵，所以有 $AA^H = E$。即 A 为酉矩阵。

13. 由于

$$A \cdot A^H = \begin{bmatrix} P \cdot P^H + B \cdot B^H & B \cdot Q^H \\ Q \cdot B^H & Q \cdot Q^H \end{bmatrix} = E,$$

可得

$$\begin{cases} Q \cdot Q^H = E_n (n \text{ 阶单位阵}) & (1) \\ Q \cdot B^H = 0 (\text{矩阵}) & (2) \end{cases}$$

由(1)式知 Q 为酉矩阵；再应用(2)式可推出 $B = 0$(矩阵)。最后再由 $P \cdot P^H = E_m$，知 P 也是酉矩阵。

14. 因 $A^H = A, B^H = B$，又

$$(AB)^H = AB \Longleftrightarrow B^H A^H = AB \text{ 即 } BA = AB$$

15. 设 A 为任一复方阵，令

$$A = X + Y \tag{1}$$

其中 X 为厄米特矩阵，Y 为反厄米特矩阵。于是可得

$$A^H = X^H + Y^H = X - Y \tag{2}$$

由(1)与(2)两式得到

$$X = \frac{A + A^H}{2}, \quad Y = \frac{A - A^H}{2}$$

16. (1) A 的特征值为 $\lambda_1 = 1, \lambda_2 = -1, \lambda_3 = -2$。相应的三个特征向量为：

$$\alpha_1 = (i, 2, i)^T, \alpha_2 = (1, 0, -1)^T, \alpha_3 = (i, -1, i)^T$$

将它们单位化得 $\varepsilon_1, \varepsilon_2, \varepsilon_3$，则 $P = (\varepsilon_1, \varepsilon_2, \varepsilon_3)$。

(2) 可求得 A 的特征值为

$$\lambda_1 = 0, \quad \lambda_2 = -\sqrt{2}, \quad \lambda_3 = \sqrt{2}$$

相应的特征向量为

$$\alpha_1 = (0, i, 1)^T, \alpha_2 = (2, i, -1)^T, \alpha_3 = (2, -i, 1)^T.$$

将它们单位化得 $\varepsilon_1, \varepsilon_2, \varepsilon_3$，则 $P = (\varepsilon_1, \varepsilon_2, \varepsilon_3)$。

17. 设 A, B 是两个 n 阶正规矩阵，如果 A 与 B 是酉等价（酉相似）的，亦即存在酉矩阵 Q，使得 $B = Q^H AQ = Q^{-1}AQ$. 因而有

$$\begin{aligned} |\lambda E - B| &= |\lambda E - Q^{-1}AQ| \\ &= |Q^{-1}(\lambda E - A)Q| \\ &= |\lambda E - A| \end{aligned}$$

即 A, B 有相同的特征多项式。

反之，若 A, B 有相同的特征多项式，因而有相同的特征值集合 $\{\lambda_1, \lambda_2, \cdots, \lambda_n\}$。由 §7 的定理 1，存在酉矩阵 Q_1 及 Q_2，使得

$$Q_1^{-1}AQ_1 = \begin{pmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{pmatrix} = Q_2^{-1}BQ_2$$

因此有

$$\begin{aligned} A &= Q_1Q_2^{-1}BQ_2Q_1^{-1} = (Q_2Q_1^{-1})^{-1}B(Q_2Q_1^{-1}) \\ &= P^{-1}BP. \end{aligned}$$

易知 $P = Q_2Q_1^{-1}$ 是酉矩阵，即 A, B 是酉相似的。

18.（证）分下列四个步骤来证：

1）由 $AB = BA$，推知 A, B 至少有一个公共的特征向量。

事实上，设 R_λ 是 A 的属于特征值 λ 的特征子空间。若 $\boldsymbol{x} \in R_\lambda$，即 $A\boldsymbol{x} = \lambda\boldsymbol{x}$. 则 $BA\boldsymbol{x} = \lambda B\boldsymbol{x}$，由 $BA = AB$，于是有 $A(B\boldsymbol{x}) = \lambda(B\boldsymbol{x})$. 即 $B\boldsymbol{x} \in R_\lambda$，从而 R_λ 是 B 的不变子空间，故在 R_λ 中存在 B 的特征向量 $\boldsymbol{y}$。显然它也是 A 的特征向量。

2）由 $AB = BA$，推知 A, B 可同时酉相似于上三角阵。即有酉阵 Q，使 $Q^H AQ$ 及 $Q^H BQ$ 均为上三角阵。

事实上，对矩阵 A 与 B 的阶 n 行归纳法。当 $n = 1$ 时结论显然成立。今设单位向量 $\boldsymbol{x}^{(1)}$ 是 A, B 的一个公共特征向量。再适当补充

$n-1$ 个单位向量 $x^{(2)},\cdots,x^{(n)}$，使 $\{x^{(1)},x^{(2)},\cdots,x^{(n)}\}$ 为标准正交基，从而 $P=(x^{(1)},x^{(2)},\cdots,x^{(n)})$ 为一酉矩阵。且有

$$Bx^{(1)}=\beta x^{(1)},BP=(\beta x^{(1)},Bx^{(2)},\cdots,Bx^{(n)}).$$

从而有

$$P^H BP=\begin{pmatrix}\beta & b\\ 0 & B_1\end{pmatrix}=B^*.$$

这里 b 是 $1\times(n-1)$ 矩阵，B_1 是 $n-1$ 阶方阵。

$$P^H AP=\begin{pmatrix}\gamma & a\\ 0 & A_1\end{pmatrix}=A^*,$$

这里 a 是 $1\times(n-1)$ 矩阵，A_1 是 $n-1$ 方阵。

由 $BA=AB$ 即有：

$$(PB^*P^H)\cdot(PA^*P^H)=(PA^*P^H)\cdot(PB^*P^H)$$

于是得 $B^*A^*=A^*B^*$。由此可推得 $B_1A_1=A_1B_1$。故由归纳法假设，存在 $n-1$ 阶酉矩阵 P_1，使得

$$P_1^H B_1 P_1=\Delta$$

为上三角矩阵，令

$$T=\begin{pmatrix}1 & 0\\ 0 & P_1\end{pmatrix},\quad Q=PT,$$

则有

$$\begin{aligned}Q^H BQ&=T^H(P^H BP)T\\ &=\begin{pmatrix}1 & 0\\ 0 & P_1^H\end{pmatrix}\begin{pmatrix}\beta & b\\ 0 & B_1\end{pmatrix}\begin{pmatrix}1 & 0\\ 0 & P_1\end{pmatrix}=\begin{pmatrix}\beta & bP_1\\ 0 & \Delta\end{pmatrix}\end{aligned}$$

这是个上三角矩阵。易证 Q 是酉矩阵，故 B 酉相似于上三角矩阵。同理，$Q^H AQ$ 也是上三角矩阵。

3) 由 $AB=BA$ 且 A,B 均为正规矩阵，则 A,B 可同时酉相似于对角形矩阵。

事实上，设 $Q^H AQ=T$（T 为上三角阵）。则 $Q^H A^H Q=T^H$ 而且

$$AA^H = QTQ^H \cdot QT^HQ^H = Q(TT^H)Q^H.$$
$$A^HA = QT^HQ^H \cdot QTQ^H = Q(T^HT)Q^H.$$

由 $AA^H = A^HA =$ (因 A 正规),可得出 $TT^H = T^HT$,从而又可推出 T 为对角形矩阵。

对 Q^HBQ 也可作同样的证明。

4) 由 A,B 可换且正规,故由 3)可设:

$$Q^HAQ = \begin{pmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{pmatrix}, \quad Q^HBQ = \begin{pmatrix} \mu_1 & & & \\ & \mu_2 & & \\ & & \ddots & \\ & & & \mu_n \end{pmatrix}.$$

于是有

$$Q^HABQ = \begin{pmatrix} \lambda_1\mu_1 & & & \\ & \lambda_2\mu_2 & & \\ & & \ddots & \\ & & & \lambda_n\mu_n \end{pmatrix}$$

应用§7 定理 1,便知 AB 为正规矩阵。

习 题 三

1. $A\boldsymbol{x} = \lambda\boldsymbol{x}, A^2\boldsymbol{x} = \lambda A\boldsymbol{x} = \lambda^2\boldsymbol{x}$,即 $E\boldsymbol{x} = \lambda^2\boldsymbol{x}$,故$(\lambda^2 - 1)\boldsymbol{x} = \theta$, $\lambda^2 = 1, \lambda = \pm 1.\ (\boldsymbol{x} \neq \theta)$.

2. $A\boldsymbol{x} = \lambda\boldsymbol{x}, A^2\boldsymbol{x} = \lambda^2\boldsymbol{x}$,一般 $A^m\boldsymbol{x} = \lambda^m\boldsymbol{x}, (\boldsymbol{x} \neq \theta)$. 所以 λ^m 是 A^m 的特征值,而 A^m 的属于特征值 λ^m 的特征向量是 $\boldsymbol{x}$。

3. 由于 $A = (a_{ij})_{n\times n}$ 的任一行的元素之和均等于 a,故下式显然成立:

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix} \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} = a \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}$$

4.(1)特征值为 7,-2。相应的特征向量为

$$\begin{pmatrix}1\\1\end{pmatrix},\quad \begin{pmatrix}-4\\5\end{pmatrix}$$

又 $P=\begin{pmatrix}1 & -4\\1 & 5\end{pmatrix}$, $P^{-1}AP=\begin{pmatrix}7 & 0\\0 & -2\end{pmatrix}$

(2)相似于对角形

$$\begin{bmatrix}1 & 0 & 0\\0 & 2 & 0\\0 & 0 & 3\end{bmatrix}$$

(3)相似于约当标准形

$$\begin{bmatrix}2 & 0 & 0\\0 & 2 & 0\\0 & 1 & 2\end{bmatrix}$$

5. 如 A 可逆,则 $BA=A^{-1}(AB)A$, 又

$$|\lambda E-BA|=|A^{-1}(\lambda E-AB)A|=|\lambda E-AB|$$

6.

(1) $\begin{bmatrix}2 & 0 & 0\\0 & 0 & 0\\0 & 0 & 0\end{bmatrix}$, (2) $\begin{bmatrix}1 & 0 & 0\\0 & i & 0\\0 & 0 & -i\end{bmatrix}$

(3) $\begin{bmatrix}-1 & 0 & 0\\0 & -1 & 0\\0 & 1 & -1\end{bmatrix}$, (4) $\begin{bmatrix}1 & 0 & 0\\1 & 1 & 0\\0 & 1 & 1\end{bmatrix}$

7. 设 $P^{-1}A^TP=J$(约当形),则 $P^TA(P^{-1})^T=J^T$,即 $Q^{-1}AQ=J^T$(上三角阵),这里 $Q=(P^{-1})^T=(P^T)^{-1}$.

8.(1) 如 λ 为 A 的任一特征值,则 λ^m 为 A^m 的特征值。若 $A^m=0$(零矩阵),则 $|\lambda E-A^m|=|\lambda E|=\lambda^n$,即 A^m 为零矩阵时 A 的特征值全为零。反之,若 A 的所有特征值均为零,则 A 的特征多项式为 $f(\lambda)=\lambda^n$.(A 是 n 阶方阵,上述 n 与此同)。故由哈密顿 —— 开来定理,有 $f(A)=0$(矩阵),即 $A^n=0$(矩阵)。

(2) 设 A 的约当标准形为

$$J=\begin{pmatrix}J_1 & & & \\ & J_2 & & \\ & & \ddots & \\ & & & J_k\end{pmatrix}$$

则由 $A^m=0$(矩阵)得知 A 的所有特征值皆为零，故 $J_1,J_2,\cdots,J_k$ 的主对角线上元素均为零，从而 J 的主对角线上元素全为零，因此 $J+E$ 为主对角线上元素全为 1 的下三角矩阵。于是有

$$|A+E|=|C^{-1}JC+E|=|C^{-1}(J+E)C|=|J+E|=1$$

9

(1) $\begin{pmatrix}\lambda & 0 \\ 0 & \lambda^3-10\lambda^2-3\lambda\end{pmatrix}$.　　(2) $\begin{bmatrix}1 & 0 & 0 \\ 0 & \lambda & 0 \\ 0 & 0 & \lambda^2+\lambda\end{bmatrix}$

(3) $\begin{bmatrix}1 & 0 & 0 \\ 0 & \lambda(\lambda+1) & 0 \\ 0 & 0 & \lambda(\lambda+1)^2\end{bmatrix}$

10. $f(\lambda)=|\lambda E-A|=\lambda^n+a_1\lambda^{n-1}+a_2\lambda^{n-2}+\cdots+(-1)^n|A|$，由 $f(A)=0$(矩阵)，得 $A^n+a_1A^{n-1}+\cdots+(-1)^n|A|E=0$(矩阵)，即 $A(A^{n-1}+a_1A^{n-2}+a_2A^{n-3}+\cdots)=(-1)^{n+1}|A|E$，故有

$$A^{-1}=\frac{1}{(-1)^{n+1}|A|}(A^{n-1}+a_1A^{n-2}+a_2A^{n-3}+\cdots)$$

11. $f(\lambda)=|\lambda E-A|=\lambda^2-6\lambda+7$，又作多项式

$$\varphi(\lambda)=2\lambda^4-12\lambda^3+19\lambda^2-29\lambda+37.$$

则 $\varphi(\lambda)=f(\lambda)(2\lambda^2+5)+\lambda+2$. 由 $f(A)=0$(矩阵)，得 $\varphi(A)=A+2E$，即

$$2A^4-12A^3+19A^2-29A+37E=\varphi(A)$$

$$=A+2E=\begin{pmatrix}3 & -1 \\ 2 & 7\end{pmatrix},$$

由此看出 $\varphi(A)$ 可逆，又

$$|\lambda E-(A+2E)| = \lambda^2 - 10\lambda + 23,$$

$$(A+2E)^2 - 10(A+2E) + 23E = 0(\text{矩阵}),$$

$$(A+2E)(A-8E) = -23E.$$

故所求逆阵为：

$$(A+2E)^{-1} = -\frac{1}{23}(A-8E)$$

12. $(E+B(E-AB)^{-1}A)\cdot(E-BA)$

$$\begin{aligned}
&= E - BA + B(E-AB)^{-1}A(E-BA)\\
&= E - BA + B(E-AB)^{-1}(A-ABA)\\
&= E - BA + B(E-AB)^{-1}(E-AB)A\\
&= E - BA + B\cdot E\cdot A\\
&= E.
\end{aligned}$$

故 $E-BA$ 可逆，且逆阵为 $E+B(E-AB)^{-1}A$.

13. $f(\lambda) = |\lambda E - A| = \lambda^3 - \lambda^2 - \lambda + 1$，由 $f(A) = 0$(矩阵)得 $A^3 = A + A^2 - E$，即在 $n=3$ 时等式 $A^n = A^{n-2} + A^2 - E$ 成立。对 $n+1$ 则有 $A^{n+1} = A\cdot A^n = A(A^{n-2}+A^2-E) = A^{n-1} + A^3 - A = A^{n-1} + (A+A^2-E) - A = A^{n-1} + A^2 - E$，又

$$\begin{aligned}
A^{100} &= A^{98} + A^2 - E\\
&= A^{96} + 2(A^2-E) = \cdots(\text{重复应用公式 49 次})\\
&= A^2 + 49(A^2-E) = 50A^2 - 49E\\
&= \begin{pmatrix} 1 & 0 & 0\\ 50 & 1 & 0\\ 50 & 0 & 1\end{pmatrix}
\end{aligned}$$

14. 因 $(A+2E)(A-E) = 0$(矩阵)，故 $\varphi(\lambda) = (\lambda+2)(\lambda-1)$ 是 A 的化零多项式，但最小多项式 $m(\lambda)$ 可整除 $\varphi(\lambda)$，故 $m(\lambda)$ 无重根，从而 A 可与对角形相似。

15. 设 $A = P^{-1}JP$，J 为约当矩阵，且

$$J=\begin{pmatrix}J_1 & & & \\ & J_2 & & \\ & & \ddots & \\ & & & J_k\end{pmatrix},\quad J_i\text{ 为第 } i \text{ 个约当块}$$

又取

$$H=\begin{pmatrix}H_1 & & & \\ & H_2 & & \\ & & \ddots & \\ & & & H_k\end{pmatrix},\text{而 } H_i=\begin{pmatrix} & & 1 \\ & \cdot^{\cdot^{\cdot}} & 1 & \\ 1 & & \end{pmatrix}\text{与 } J_i \text{ 同阶。}$$

则 $H_i^2=E$，即 $H_i^{-1}=H_i$，且由计算得 $J=H^{-1}J^TH$。故 $A=P^{-1}JP=P^{-1}H^{-1}J^THP=(HP)^{-1}[(P^{-1})^TA^TP^T](HP)=(P^THP)^{-1}A^T\cdot(P^THP)=C^{-1}A^TC$，其中 $C=P^THP$，又 C 是可逆的，对称的。

设 $D=AC^{-1}$，则有 $A=DC$，而 $D^T=(C^{-1})^TA^T=(C^T)^{-1}C^TA(C^{-1})^T=AC^{-1}=D$，即 D 也是对称的。

16.(1)由 §5 定理 3，因多项式矩阵

$$\begin{pmatrix}D(\lambda)\\ N(\lambda)\end{pmatrix}$$

的史密斯标准形为

$$\begin{pmatrix}1 & 0\\ 0 & 1\\ 0 & 3\lambda+1\end{pmatrix},$$

故 $D(\lambda)$，$N(\lambda)$ 不是右互质的。

(2)因多项式矩阵

$$\begin{pmatrix}D(\lambda)\\ N(\lambda)\end{pmatrix}$$

的史密斯标准形为

$$\begin{pmatrix}1 & 0\\ 0 & 1\\ 0 & 0\\ 0 & 0\end{pmatrix}$$

故 $D(\lambda)$，$N(\lambda)$ 是右互质的。

17. $\deg|M(\lambda)| = 6 = kc_1 + kc_2 + kc_3 = 3 + 1 + 2$，故由定义知 $M(\lambda)$ 是列既约的。但 $kr_1 + kr_2 + kr_3 = 3 + 3 + 2 \neq \deg|M(\lambda)| = 6$，故 $M(\lambda)$ 不是行既约的。

18. 史密斯-麦克米伦标准形为

$$M(\lambda) = \begin{bmatrix} \dfrac{\lambda}{(\lambda+1)^2(\lambda+2)^2} & 0 \\ 0 & \dfrac{\lambda^2}{\lambda+2} \end{bmatrix}$$

又 $G(\lambda)$ 的一个右分解为：

$$\begin{pmatrix} \lambda & \lambda \\ -\lambda(\lambda+1)^2 & -\lambda \end{pmatrix} \begin{bmatrix} (\lambda+1)^2(\lambda+2)^2 & 0 \\ 0 & (\lambda+2)^2 \end{bmatrix}^{-1}$$

或

$$\begin{pmatrix} \lambda & 0 \\ -\lambda & \lambda^2 \end{pmatrix} \begin{pmatrix} 0 & -(\lambda+1)^2(\lambda+2)^2 \\ (\lambda+2)^2 & \lambda+2 \end{pmatrix}^{-1}$$

习 题 四

1. (1) $\|\theta\| = \|0 \cdot \boldsymbol{x}\| = 0 \cdot \|\boldsymbol{x}\| = 0.$

(2) $\left\|\dfrac{\boldsymbol{x}}{\|\boldsymbol{x}\|}\right\| = \dfrac{1}{\|\boldsymbol{x}\|}\|\boldsymbol{x}\| = 1.$

(3) $\|-\boldsymbol{x}\| = |(-1)| \cdot \|\boldsymbol{x}\| = \|\boldsymbol{x}\|.$

(4) 因 $\|\boldsymbol{x}\| = \|\boldsymbol{x}-\boldsymbol{y}+\boldsymbol{y}\| \leqslant \|\boldsymbol{x}-\boldsymbol{y}\| + \|\boldsymbol{y}\|$，从而有 $\|\boldsymbol{x}\| - \|\boldsymbol{y}\| \leqslant \|\boldsymbol{x}-\boldsymbol{y}\|$. 另方面又有

$$\begin{aligned} \|\boldsymbol{x}-\boldsymbol{y}\| &= \|\boldsymbol{y}-\boldsymbol{x}\| \geqslant \|\boldsymbol{y}\| - \|\boldsymbol{x}\| \\ &= -(\|\boldsymbol{x}\| - \|\boldsymbol{y}\|), \end{aligned}$$

所以又有 $\|\boldsymbol{x}\| - \|\boldsymbol{y}\| \geqslant - \|\boldsymbol{x}-\boldsymbol{y}\|$.

由实数绝对值不等式性质(u，a 是实数)

$$-a \leqslant u \leqslant a \Longleftrightarrow |u| \leqslant a,$$

便有不等式

$$|\|x\| - \|y\|| \leqslant \|x-y\|.$$

2.(1)因为

$$\begin{aligned}\|\boldsymbol{x}\|_1^2 &= (|\xi_1| + \cdots + |\xi_n|)^2 \\ &= |\xi_1|^2 + |\xi_2|^2 + \cdots + |\xi_n|^2 \\ &\quad + 2|\xi_1|\cdot|\xi_2| + \cdots + 2|\xi_{n-1}|\cdot|\xi_n| \\ &\geqslant |\xi_1|^2 + |\xi_2|^2 + \cdots + |\xi_n|^2 = \|\boldsymbol{x}\|_2^2\end{aligned}$$

所以 $\|\boldsymbol{x}\|_2 \leqslant \|\boldsymbol{x}\|_1$。又因

$$\begin{aligned}&n\|\boldsymbol{x}\|_2^2 - \|\boldsymbol{x}\|_1^2 \\ &= n(|\xi_1|^2 + \cdots + |\xi_n|^2) - (|\xi_1| + \cdots + |\xi_n|)^2 \\ &= \sum_{1\leqslant i<j\leqslant n}(|\xi_i| - |\xi_j|)^2 \geqslant 0\end{aligned}$$

因此又得 $\|\boldsymbol{x}\|_1 \leqslant \sqrt{n}\|\boldsymbol{x}\|_2$。

(2) 设 $\max\limits_{1\leqslant i\leqslant n}|\xi_i| = |\xi_k|$,则 $\|\boldsymbol{x}\|_\infty = |\xi_k|$. 另方面又有:

$$\|\boldsymbol{x}\|_1 = |\xi_1| + \cdots + |\xi_n| \geqslant |\xi_k| = \|\boldsymbol{x}\|_\infty,$$

$$\|\boldsymbol{x}\|_1 = |\xi_1| + \cdots + |\xi_n| \leqslant n|\xi_k| = n\|\boldsymbol{x}\|_\infty.$$

(3)由下不等式便可得证($|\xi_k|$ 见(2)):

$$\begin{aligned}\|\boldsymbol{x}\|_\infty^2 = |\xi_k|^2 &\leqslant \sum|\xi_i|^2 = \|\boldsymbol{x}\|_2^2 = \sum|\xi_i|^2 \\ &\leqslant n|\xi_k|^2 = n\|\boldsymbol{x}\|_\infty^2.\end{aligned}$$

3. 柯西不等式 $|\boldsymbol{x}^T\boldsymbol{y}| = |(\boldsymbol{x},\boldsymbol{y})| \leqslant \|\boldsymbol{x}\|_2\cdot\|\boldsymbol{y}\|_2$ 当且只当 $\boldsymbol{x},\boldsymbol{y}$ 线性相关时等号成立,即

$$|(\boldsymbol{x},\boldsymbol{y})| = \|\boldsymbol{x}\|_2\cdot\|\boldsymbol{y}\|_2.$$

又 $\boldsymbol{x}^T\boldsymbol{y} \geqslant 0$ 时,$|\boldsymbol{x}^T,\boldsymbol{y}| = |(\boldsymbol{x},\boldsymbol{y})| = (\boldsymbol{x},\boldsymbol{y})$.

4.(1)T 为正交矩阵,$T'T = E$,这里 T' 表示 T 的转置矩阵。而 E 的特征值为1。故由定义,有(注意 $\lambda_{T'T}$ 表示 $T'T = E$ 的最大特征值):

$$\|T\|_2 = \sqrt{\lambda_{T'T}} = \sqrt{\lambda_E} = 1$$

(2) $\| TA \|_2 = \sqrt{\lambda_{(TA)'(TA)}} = \sqrt{\lambda_{A'A}} = \| A \|_2$.

5. $\| A \|_2 = \sqrt{\lambda_{A^H A}}$，矩阵 $A^H A$ 是个特殊的厄米特矩阵，其特征值是非负实数（见 148 页的注），不妨设这些特征值为：

$$\lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_n \geqslant 0$$

于是得 $\| A \|_2 = \sqrt{\lambda_1}$，且

$$\| A \|_F = \sqrt{\operatorname{tr}(A^H A)} = \sqrt{\sum_1^n \lambda_i} \geqslant \sqrt{\lambda_1} = \| A \|_2$$

$$\| A \|_F = \sqrt{\sum_1^n \lambda_i} \leqslant \sqrt{n\lambda_1} = \sqrt{n}\, \| A \|_2.$$

注：n 阶矩阵 $C = (c_{ij})_{n\times n}$ 的迹 $\operatorname{tr}C$ 定义为

$$\operatorname{tr}C = c_{11} + c_{22} + \cdots + c_{nn}.\text{（见 78 页）}$$

而特征多项式

$$|\lambda E - C| = \lambda^n + a_1\lambda^{n-1} + \cdots + a_n = \varphi(\lambda)$$

中系数 $a_1 = -\operatorname{tr}C$. 又根据多项式理论，有

$$\lambda_1 + \lambda_2 + \cdots + \lambda_n = -a_1.$$

这里 λ_1 ，λ_2 ，$\cdots$ ，λ_n 为 $\varphi(\lambda)$ 的 n 个根。从而 $\sum_1^n \lambda_i = \operatorname{tr}C$. 关于特征多项式的系数的计算公式，以及多项式的根与系数的关系式，其推导都是困难的，读者不必去证。

6. $e^A = \begin{pmatrix} e & 2e \\ 0 & e \end{pmatrix}$

7

(1) $e^{At} = \begin{pmatrix} 2e^{-t} - e^{-2t} & e^{-t} - e^{-2t} \\ -2e^{-t} + 2e^{-2t} & -e^{-t} + 2e^{-2t} \end{pmatrix}$

(2) $-\dfrac{e^t}{6}\begin{bmatrix} -3 & 5 & -2 \\ 3 & -5 & 2 \\ 3 & -5 & 2 \end{bmatrix} + \dfrac{e^{-2t}}{15}\begin{bmatrix} 0 & 11 & -11 \\ 0 & 1 & -1 \\ 0 & -14 & 14 \end{bmatrix}$

$$+\frac{e^{3t}}{10}\begin{bmatrix}5 & 1 & 4\\ 5 & 1 & 4\\ 5 & 1 & 4\end{bmatrix}.$$

(3). $$\begin{bmatrix}2t^2+2t+1 & 2(t+1) & \frac{1}{2}t^2\\ -4t^2 & -4t^2+2t+1 & -t^2-t\\ 8t(t-5) & 4t(2t-3) & 2t^2-12t+1\end{bmatrix}e^{-2t}.$$

(4) $$\begin{bmatrix}1 & -5 & 1\\ 0 & 0 & 0\\ 0 & 2 & 1\end{bmatrix}e^{-2t}+\begin{bmatrix}0 & 6 & 3\\ 0 & 0 & 0\\ 0 & 0 & 0\end{bmatrix}te^{-2t}+\begin{bmatrix}0 & 5 & 0\\ 0 & 1 & 0\\ 0 & -2 & 0\end{bmatrix}e^{-3t}.$$

8. $$e^A=\begin{bmatrix}1 & \frac{1}{2}(1-e^{-2})\\ 0 & e^{-2}\end{bmatrix}.\quad \sin A=\begin{bmatrix}0 & \frac{1}{2}\sin 2\\ 0 & -\sin 2\end{bmatrix}$$

$$\cos At=\begin{bmatrix}1 & \frac{1}{2}(1-\cos 2t)\\ 0 & \cos 2t\end{bmatrix}.$$

9. $$\frac{d}{dt}A(t)=\begin{pmatrix}-\sin t & \cos t\\ -\cos t & -\sin t\end{pmatrix}.$$

$$\frac{d}{dt}A^{-1}(t)=\begin{pmatrix}-\sin t & -\cos t\\ \cos t & -\sin t\end{pmatrix}.$$

$$\frac{d}{dt}|A(t)|=0.\quad \left|\frac{d}{dt}A(t)\right|=1.$$

10. $$\int A(t)dt=\begin{bmatrix}\frac{1}{2}e^{2t}+c_{11} & (t-1)e^t+c_{12} & t+c_{13}\\ -e^{-t}+c_{21} & e^{2t}+c_{22} & c_{23}\\ \frac{3}{2}t^2+c_{31} & c_{32} & c_{33}\end{bmatrix}$$

$$\int_0^1 A(t)dt=\begin{bmatrix}\frac{1}{2}(e^2-1) & 1 & 1\\ 1-e^{-1} & e^2-1 & 0\\ \frac{3}{2} & 0 & 0\end{bmatrix}$$

11. (1) 首先易证矩阵序列与级数的下述性质：

1) 若矩阵序列 $\{A_m\} \to A$，则 $\{A_m^T\} \to A^T$，$\{\bar{A}_m\} \to \bar{A}$；

2) 若矩阵级数 $\sum_0^\infty C_m A^m$ 收敛，则 $\left\{\sum_0^\infty C_m A^m\right\}^T = \sum_0^\infty C_m (A^T)^m$.

由此即得

$$(e^A)^T = \left\{\sum_0^\infty \frac{A_m}{m!}\right\}^T = \sum_0^\infty \frac{(A^T)^m}{m!} = e^{A^T}$$

当 A 是实反对称矩阵，则

$$(e^A) \cdot (e^A)^T = e^A \cdot e^{A^T} = e^A \cdot e^{-A} = e^0 = E.$$

即 e^A 为正交阵。

(3) 若 A 为厄米特矩阵，即 $A^H = A$，则

$$\begin{aligned}(e^{iA})^H \cdot e^{iA} &= e^{(iA)^H} \cdot e^{iA} = e^{-iA^H} \cdot e^{iA} \\ &= e^{-iA + iA} = e^0 = E,\end{aligned}$$

故 e^{iA} 为酉矩阵。

12. $X(t) = \begin{bmatrix} \dfrac{2}{\sqrt{3}}\sin\sqrt{3}t \\ \cos\sqrt{3}t + \dfrac{1}{\sqrt{3}}\sin\sqrt{3}t \end{bmatrix}$.

提示：

$A = \begin{pmatrix} -1 & 2 \\ -2 & 1 \end{pmatrix}$的特征值为 $\lambda = \dfrac{1 \pm \sqrt{3}i}{2}$，相应的两个特征向量为

$$\alpha = \begin{bmatrix} 1 \\ \dfrac{1+\sqrt{3}i}{2} \end{bmatrix}, \quad \beta = \begin{bmatrix} 1 \\ \dfrac{1-\sqrt{3}i}{2} \end{bmatrix}$$

作矩阵

$$P = \begin{bmatrix} 1 & 1 \\ \dfrac{1+\sqrt{3}i}{2} & \dfrac{1-\sqrt{3}i}{2} \end{bmatrix}$$

则(利用尤拉公式 $e^{ix}=\cos x+i\sin x$;当 x 为实数):

$$e^{At}=P\begin{bmatrix}e^{\sqrt{3}it} & 0\\ 0 & e^{-\sqrt{3}it}\end{bmatrix}P^{-1}$$

$$=\begin{bmatrix}\cos\sqrt{3}t-\dfrac{1}{\sqrt{3}}\sin\sqrt{3}t & \dfrac{2}{\sqrt{3}}\sin\sqrt{3}t\\ -\dfrac{2}{\sqrt{3}}\sin\sqrt{3}t & \cos\sqrt{3}t+\dfrac{1}{\sqrt{3}}\sin\sqrt{3}t\end{bmatrix}.$$

$$X(t)=e^{At}\begin{pmatrix}0\\1\end{pmatrix}.$$

13. $X(t)=\begin{pmatrix}\sin 5t\\ \cos 5t\end{pmatrix}e^{3t}+\begin{pmatrix}a(t)\\ b(t)\end{pmatrix}e^{3t}$,其中:

$$a(t)=\left(\frac{e^{-4t}}{41}(-4\cos 5t+5\sin 5t)+\frac{4}{41}\right)\cos 5t$$
$$+\left(\frac{e^{-4t}}{41}(-4\sin 5t-5\cos 5t)+\frac{5}{41}\right)\sin 5t,$$

$$b(t)=-\left(\frac{e^{-4t}}{41}(-4\cos 5t+5\sin 5t)+\frac{4}{41}\right)\sin 5t$$
$$+\left(\frac{e^{-4t}}{41}(-4\sin 5t-5\cos 5t)+\frac{5}{41}\right)\cos 5t.$$

提示:

设 $A=\begin{pmatrix}3 & 5\\ -5 & 3\end{pmatrix}$,则 A 的特征值为 $\lambda=3\pm 5i$。对应的两个线性无关特征向量为:

$$\alpha=(1,i)^T,\beta=(i,1)^T.$$

以它们为列向量,构成可逆矩阵

$$P=\begin{pmatrix}1 & i\\ i & 1\end{pmatrix}.$$

从而有

$$e^{At}=\begin{pmatrix}1 & i\\ i & 1\end{pmatrix}\begin{pmatrix}e^{(3+5i)t} & 0\\ 0 & e^{(3-5i)t}\end{pmatrix}\cdot\frac{1}{2}\begin{pmatrix}1 & -i\\ -i & 1\end{pmatrix}$$

$$=\begin{pmatrix} \cos 5t & \sin 5t \\ -\sin 5t & \cos 5t \end{pmatrix} e^{3t}.$$

$$X(t)=e^{At}\begin{pmatrix}0\\1\end{pmatrix}+\int_0^t e^{A(t-\tau)}\begin{pmatrix}e^{-\tau}\\0\end{pmatrix}d\tau$$

$$=e^{3t}\begin{pmatrix}\sin 5t\\ \cos 5t\end{pmatrix}+e^{3t}\int_0^t\begin{pmatrix}\cos 5t\cos 5\tau+\sin 5t\sin 5\tau\\ -\sin 5t\cos 5\tau+\cos 5t\sin 5\tau\end{pmatrix}e^{-4\tau}d\tau$$

习 题 五

1. 由推论 1 得

$$|\lambda_i|\leqslant 2.7,\ |R_e(\lambda_i)|\leqslant 2.7,\ |I_m(\lambda_i)|\leqslant 0.165$$

由推论 2 得 $|I_m(\lambda_i)|\leqslant 0.09526,(i=1,2,3)$

2. A 的四个特征值都落在下列的四个盖尔圆的并集内：

(1) $|z-1|\leqslant 1$；(2) $|z-1.5|\leqslant 1.5$；(3) $|z-5|\leqslant 1$；(4) $|z-5i|\leqslant 1$.

3. $\rho(A)\leqslant \|A\|_\infty=\max\limits_{1\leqslant i\leqslant 4}\sum\limits_{j=1}^{4}|a_{ij}|=\max\left\{1,1,1,\frac{6}{7}\right\}=1.$

4. 由圆盘定理 2，在由两个外切圆构成的连通部分里，矩阵 A 有且只有两个特征值。如果在每个圆上有二个特征值，则在连通部分里，A 就至少有 3 个特征值，这与前述结论矛盾。

5. 设 $QA\boldsymbol{x}=\mu\boldsymbol{x}\quad(\boldsymbol{x}\neq\theta)$. 则有

$$\boldsymbol{x}^H A^H Q^H=\bar{\mu}\boldsymbol{x}^H.$$

$$\boldsymbol{x}^H A^H Q^H\cdot QA\boldsymbol{x}=\bar{\mu}\mu\boldsymbol{x}^H\boldsymbol{x},$$

$$(A\boldsymbol{x})^H(A\boldsymbol{x})=\boldsymbol{x}^H A^H A\boldsymbol{x}=|\mu|^2|\boldsymbol{x}|^2.\ (\text{因 } Q^HQ=E)$$

但是

$$A\boldsymbol{x}=\begin{bmatrix}a_1 & & & \\ & a_2 & & \\ & & \ddots & \\ & & & a_n\end{bmatrix}\begin{bmatrix}x_1\\x_2\\ \vdots\\ x_n\end{bmatrix}=\begin{bmatrix}a_1x_1\\a_2x_2\\ \vdots\\ a_nx_n\end{bmatrix},$$

于是得：

$$(A\boldsymbol{x})^H(A\boldsymbol{x}) = (\bar{a}_1\bar{x}_1, \bar{a}_2\bar{x}_2, \cdots, \bar{a}_n\bar{x}_n)\begin{pmatrix} a_1x_1 \\ a_2x_2 \\ \vdots \\ a_nx_n \end{pmatrix}$$

$$= \sum_{i=1}^{n} |a_i|^2 |x_i|^2.$$

因此，

$$\sum_{i=1}^{n} |a_i|^2 \cdot |x_i|^2 = |\mu|^2 \cdot \sum_{i=1}^{n} |x_i|^2,$$

令 $\min\limits_{1\leqslant i\leqslant n} |a_i| = L, \max\limits_{1\leqslant i\leqslant n} |a_i| = M$, 则有

$$L = \sqrt{\frac{L^2 \sum |x_i|^2}{\sum |x_i|^2}} \leqslant |\mu| = \sqrt{\frac{\sum |a_i|^2 |x_i|^2}{\sum |x_i|^2}}$$

$$\leqslant \sqrt{\frac{M^2 \sum |x_i|^2}{\sum |x_i|^2}} = M$$

6.(1)在定理 1(§5)的四个条件中，可以看到 A 与 G(即 A^+)处于对等的地位，即 G 是 A 的广义逆 A^+(且唯一)，则 A 也是 $G(=A^+)$ 的广义逆，即 $G^+ = A$，亦即 $(A^+)^+ = A$。

(2)$(AA^+)^2 = AA^+ \cdot AA^+ = (AGA)A^+ = AA^+$.

(3) 设 A 有奇异值分解：

$$A = P\begin{pmatrix} D & 0 \\ 0 & 0 \end{pmatrix}Q^H$$

则

$$A^H = Q\begin{pmatrix} D^H & 0 \\ 0 & 0 \end{pmatrix}P^H,$$

$$(A^H)^+ = P\begin{pmatrix} (D^H)^{-1} & 0 \\ 0 & 0 \end{pmatrix}Q^H \qquad \text{(见 215 页)}$$

$$(A^{+})^{H} = P\begin{pmatrix}(Q^{-1})^{H} & 0\\ 0 & 0\end{pmatrix}Q^{H}$$

由于 $(D^{H})^{-1} = (Q^{-1})^{H}$，故 $(A^{H})^{+} = (A^{+})^{H}$.

7. (1). $A^{+} = \frac{1}{3}\begin{bmatrix} 2 & -3\\ 1 & 0\\ -1 & 3\end{bmatrix}$.

(2) $A^{+} = \frac{1}{3}\begin{pmatrix} -2i & 1 & -i\\ 1 & -i & 2\end{pmatrix}$.

8. $A^{+} = \frac{1}{8}\begin{bmatrix} 2 & -2 & 2 & 2\\ -1 & 3 & -1 & 1\\ 1 & -3 & 1 & -1\end{bmatrix}$.

9. 若 $A\boldsymbol{x} = b$ 有解 $\boldsymbol{x}_0$，则由 M—P 广义逆的条件(1)，即 $AA^{+}A = A$，得 $AA^{+}A\boldsymbol{x}_0 = b$，即 $AA^{+}\boldsymbol{b} = \boldsymbol{b}$；反之，若 $AA^{+}\boldsymbol{b} = \boldsymbol{b}$，则 $\boldsymbol{x}_0 = A^{+}\boldsymbol{b}$ 便是 $A\boldsymbol{x} = \boldsymbol{b}$ 的一个解。

10. 由 §4 的定理 2 可见矩阵 A 的广义逆矩阵 G 具有如下形式

$$G = Q\begin{pmatrix}E_r & X\\ Y & Z\end{pmatrix}P$$

这里，$r = 1$ 是 A 的秩，E_1 是一阶单位矩阵，又 X, Y, Z 分别是 1×2，1×1，1×2 矩阵，设为

$$X = [x_1, x_2], \quad Y = [y], \quad Z = [z_1, z_2]$$

又 3 阶可逆矩阵 P 与 2 阶可逆矩阵 Q，按上述定理 2 满足条件

$$PAQ = \begin{pmatrix}E_1 & \\ & 0\end{pmatrix} \qquad (*)$$

因而只须用初等行变换与列变换化矩阵 A 为($*$)式右边的形状，就可求得 P 与 Q，它们就是一些初等矩阵的积。我们显然有：

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -2 & 0 & 1 \end{bmatrix}\begin{bmatrix} 1 & 2 \\ 0 & 0 \\ 2 & 4 \end{bmatrix}\begin{pmatrix} 1 & -2 \\ 0 & 1 \end{pmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}$$

从而

$$\begin{aligned} G &= Q\begin{pmatrix} E_1 & X \\ Y & Z \end{pmatrix}P \\ &= \begin{pmatrix} 1 & -2 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} 1 & x_1 & x_2 \\ y & z_1 & z_2 \end{pmatrix}\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -2 & 0 & 1 \end{bmatrix} \\ &= \begin{pmatrix} 1 - y - 2(x_2 - z_2) & x_1 - z_1 & x_2 - z_2 \\ y - 2z_2 & z_1 & z_2 \end{pmatrix} \end{aligned}$$

取 $x_1 = z_1 = 1, x_2 = z_2 = y = 0$ 可得一个广义逆为：

$$A^- = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}$$

从而得方程 $\boldsymbol{Ax} = \boldsymbol{b}$ 的最小范数解

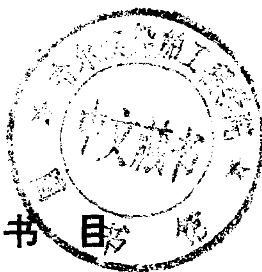
$$A^- \boldsymbol{b} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}\begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

又由第 8 题的公式可求出广义逆 A^+：

$$\begin{aligned} A^+ &= \lim_{\varepsilon \to 0}\begin{pmatrix} 1 & 0 & 2 \\ 2 & 0 & 4 \end{pmatrix}\begin{bmatrix} 5+\varepsilon^2 & 0 & 10 \\ 0 & \varepsilon^2 & 0 \\ 10 & 0 & 20+\varepsilon^2 \end{bmatrix}^{-1} \\ &= \frac{1}{25}\begin{pmatrix} 1 & 0 & 2 \\ 2 & 0 & 4 \end{pmatrix} \end{aligned}$$

而极小最小二乘解为：

$$\boldsymbol{x} = A^+ \boldsymbol{b} = \begin{bmatrix} \frac{1}{5} \\ \frac{2}{5} \end{bmatrix}.$$



参 考 书 目

〔1〕Φ.P. 甘特马赫尔,矩阵论,高等教育出版社,1955.

〔2〕И.M. 盖尔冯德,线性代数学,高等教育出版社,1957.

〔3〕北京大学数学系,高等代数,人民教育出版社,1978.

〔4〕蒋尔雄等,线性代数,人民教育出版社,1979.

〔5〕黄　琳,系统与控制理论中的线性代数,科学出版社,1984.

〔6〕陈公宁,矩阵理论与应用,高等教育出版社,1990.

〔7〕丁学仁等,工程中的矩阵理论,天津大学出版社,1985.

〔8〕R.A. 合恩等,矩阵分析,天津大学出版社,1989.

〔9〕李　乔,矩阵论八讲,上海科学技术出版社,1988.

〔10〕蒋正新等,矩阵理论及其应用,北京航空学院出版社,1988.

〔11〕R. Bellman, Introduction to Matrix Analysis, McGraw—Hill, N. Y., 1970.

〔12〕A. S. Deif, Advanoed Matrix Theory for Scientists and Engineers, Abacus Press, 1982.

〔13〕郑大钟,线性系统理论,清华大学出版社,1990.

〔14〕何　钺,现代控制理论基础,机械工业出版社,1987.

〔15〕蔡大用,数值代数,清华大学出版社,1987.

〔16〕И.C,索明斯基等,高等代数习题集,高等教育出版社,1956.

〔17〕И.B. 普罗斯库烈柯夫,线性代数习题集,人民教育出版社,1981.

〔18〕杨子胥,高等代数习题解,山东科学技术出版社,1982.

〔19〕C. R. Rao and S. K. Mitra, Generalized Inverse of Matrices and its Applications, JOHN WILEY & SONS, INC, 1971.